

Stein's Method for Multivariate Extremes

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von

Anne Feidt

aus

Luxemburg

Promotionskomitee

Prof. Dr. Erwin Bolthausen (Vorsitz)
Prof. Dr. Andrew Barbour

Zürich, 2013

Abstract

We apply the Stein-Chen method to problems from extreme value theory. On the one hand, the Stein-Chen method for Poisson approximation allows us to obtain bounds on the Kolmogorov distance between the law of the maximum of i.i.d. random variables, following certain well known distributions, and an extreme value distribution. On the other hand, we introduce marked point processes of exceedances (MPPE's) whose i.i.d. marks can be either univariate or multivariate. We use the Stein-Chen method for Poisson process approximation to determine bounds on the error of the approximation, in some appropriate probability metric, of the law of the MPPE by that of a Poisson process. The Poisson process that we approximate by has intensity measure equal to that of the MPPE. In some cases, this intensity measure is difficult to work with, or varies with the sample size; we then approximate by a further easier-to-use Poisson process and estimate the error of this additional approximation.

Zusammenfassung

Wir wenden die Stein-Chen Methode auf Fragestellungen der Extremwerttheorie an. Einerseits erlaubt uns die Stein-Chen Methode für Poisson-Approximation die Kolmogorow-Distanz zwischen der Verteilung des Maximums von i.i.d. Zufallsvariablen und einer Extremwertverteilung nach oben abzuschätzen. Andererseits führen wir markierte Punktprozesse von Grenwertüberschreitungen (MPPE genannt) ein, deren Markierungen sowohl univariat als auch multivariat sein können. Wir verwenden die Stein-Chen Methode zur Approximation durch Punktprozesse, um den Fehler der Approximation der Verteilung eines MPPE durch die Verteilung eines Poissonprozesses in einer geeigneten Metrik abzuschätzen. Das Intensitätsmass des approximierenden Poissonprozesses ist dasselbe wie das des MPPEs. In manchen Fällen ist dieses Intensitätsmass schwierig anwendbar; wir approximieren dann durch einen weiteren, einfacheren Poissonprozess und bestimmen den Fehler dieser zusätzlichen Approximation.

Acknowledgements

Special thanks go to Prof. A. D. Barbour for taking me on as PhD student and introducing me to Stein's method. I am most grateful for the time, patience and care he devoted to my benefit. His advice and continued support were essential toward the completion of this thesis. I would also like to thank Prof. E. Bolthausen for taking over my project and Prof. J. Hüsler for acting as external referee. Further thanks go to Prof. J. Nešlehová and Prof. P. Embrechts for first introducing me to the study of extremes.

I am grateful to my officemates, Philipp Thomann, Dominik Tasnady, Rajat Hazra and Martin Gallauer for their support, both mathematical and non-mathematical, throughout the years. I likewise thank Anna Pósfai, as well as my fellow assistants at the University of Zurich, in particular Alessandra Cipriani, Kim Dang, Noemi Kurt, Erich Baur, Felix Fontein, Dominik Heinzmann and Felix Rubin.

Moreover, I would like to thank Caroline, Françoise, Lynn, Martha, Brian, Claude, Georges and Josy for being my Luxembourgish family in Zurich. Finally, special thanks go to my parents for their constant support, and to Sigi, for everything.

Contents

1	Introduction	9
2	The Stein-Chen method	17
2.1	Poisson approximation of the binomial distribution	17
2.2	Distances between probability measures	21
2.3	Le Cam's result for Poisson approximation	24
2.4	The Stein-Chen method for Poisson approximation	27
2.4.1	Construction of the Stein equation and smoothness estimates	30
2.4.2	Independent indicator variables	31
2.4.3	Dependent indicator variables – The local approach	33
2.4.4	The antisymmetric function approach	35
2.4.5	Immigration-death processes and the generator interpretation	37
2.5	Poisson process approximation in the total variation distance	40
2.5.1	Point process	41
2.5.2	Poisson process	43
2.5.3	Palm processes and Janossy densities	44
2.5.4	Approximation of point processes – the generator interpretation	46
2.5.5	Approximation of point processes with i.i.d. marks – Michel's argument	59
2.6	Improved rates for Poisson process approximation using the d_2 -distance	61
3	Poisson and Poisson process approximation for univariate extremes	77
3.1	Poisson approximation for the number of extreme points and maxima of random variables	77
3.1.1	Maxima of continuous random variables	83
3.1.2	The problem for maxima of discrete random variables	95

3.1.3	Maxima of discrete random variables	99
3.2	Poisson process approximation for MPPE's	105
3.2.1	Application to MPPE's with continuous marks	109
3.2.2	Application to MPPE's with geometric marks	116
3.2.3	Remarks on the choice of the point process and its approximation by a Poisson process	121
4	Poisson process approximation for multivariate extremes	123
4.1	Copulas and tail dependence	128
4.1.1	Definition and properties of copulas	128
4.1.2	Fréchet-Hoeffding bounds	130
4.1.3	Sklar's Theorem	131
4.1.4	Survival copulas	132
4.1.5	Absolutely continuous and singular components of copulas	134
4.1.6	Coefficients of tail dependence	136
4.2	Poisson process approximation for MPPE's with multivariate marks	137
4.2.1	Joint threshold exceedances	138
4.2.2	Single-component threshold exceedances	142
4.2.3	An example of hidden dependence for joint threshold exceedances	148
4.3	Archimedean copulas with upper tail dependence	150
4.3.1	An asymptotic result	153
4.3.2	Approximation in d_{TV} by a Poisson process	155
4.3.3	Examples	167
4.3.4	Higher dimensions	174
4.4	MPPE's with bivariate Marshall-Olkin geometric marks	178
4.4.1	The bivariate Marshall-Olkin geometric distribution	179
4.4.2	Approximation in d_{TV} by a Poisson process on a lattice	181
4.4.3	Construction of a "continuous" intensity function	183
4.4.4	Assumptions on the distributional parameters	188
4.4.5	Approximation in d_2 by a Poisson process with continuous intensity	191
4.4.6	Approximation in d_2 and d_{TV} by a Poisson process independent of n	199
4.4.7	Final bound in the d_2 -distance	204

1 Introduction

Rare events of unexpected magnitude and momentous impact are playing an ever larger role all over the globe. In addition to ever-present threats of natural disasters such as earthquakes, volcanic eruptions or tsunamis, the last few decades have seen more and more natural catastrophes that may have been induced by a changing climate due to global warming. Prime examples are devastating cyclones like North Atlantic Hurricanes Andrew in 1992, Katrina in 2005, Sandy in 2012, or like North Indian Ocean Cyclone Aila in 2009. Other natural disasters are extreme rainfalls and floods such as the 2005 European or 2010 Pakistan floods, as well as heat waves as experienced in Europe in 2003 or in North America in 2012. Further examples of extreme events are given by environmental disasters such as the 2010 Gulf of Mexico oil spill or the 2011 Fukushima Daiichi nuclear disaster. Such events translate directly into extreme financial losses as damages need to be repaired. Extremely high financial losses can of course also arise from other causes, most prominently from stock market crashes such as the Wall Street Crash of 1929 or the 2008 financial crisis. It is clear that it is of paramount importance to be able to model and predict the size and frequency of such events in order to both establish adequate emergency measures to warrant the safety of people in danger from looming disasters, and to correctly determine minimum capital requirements of financial institutions such as insurance and reinsurance companies to prevent bankruptcy and ensure coverage of damages.

The study of extremes has become a well-established field in probability theory and statistics. The classical approach in Extreme Value Theory (EVT) is to model extremes as the maximum or minimum values of a sample of n independent and identically distributed (i.i.d.) random variables X_1, \dots, X_n . Since $\min_{1 \leq i \leq n} X_i = -\max_{1 \leq i \leq n} (-X_i)$, it suffices to study maxima. A first major contribution by Fisher and Tippet (1928) assures that if there exists an affine transformation under which the maximum of i.i.d. random variables converges in distribution to a non-degenerate distribution function H , then H is an extreme value distribution, that is, either a Fréchet, Weibull or Gumbel distribution. Another approach in EVT, among others, is to consider point

processes, for instance, point processes of exceedances that indicate the points exceeding a threshold, or marked point processes that are random configurations of points in space. Such point processes have been treated in detail in, for instance Leadbetter et al. (1983) and Resnick (1987), respectively. The main result is that under certain conditions they asymptotically, for $n \rightarrow \infty$, behave like Poisson processes.

Extreme value distributions and Poisson processes are generally used as approximations for the distributions of maxima and of point processes of exceedances, respectively, for any finite sample size n . However, such approximations only make sense if we have some estimate of the errors involved in terms of n . Underlying both kinds of approximations is Poisson approximation, since the number of those among the n random variables of the sample that exceed a threshold is binomial and may thus be approximated by a Poisson distribution, provided that the probability of a threshold exceedance is small. It thus makes sense to attempt to determine the accuracy of the above approximations in some probability metric by way of Stein's method for Poisson approximation.

Stein's method is a way to determine explicit bounds on the error involved in approximating one probability distribution by another. It was first introduced by Stein (1972) for approximation by the normal distribution. The method is however applicable to approximations by other distributions. Stein's method for Poisson approximation was developed by Chen (1975a,b) and is therefore often called the Stein-Chen method. The method compares expectations of test functions under the two distributions, with the choice of test functions determining the probability metric in which the error will be expressed. For example, with the choice of indicator functions of measurable subsets of the state space, the difference between the expectations gives rise to the total variation distance. It is linked to an identity which characterises the approximating distribution via an equation, called the *Stein equation*, whose solution needs to be determined in terms of the test functions. Instead of bounding the difference of the expectations, it turns out to be easier to bound the other side of the equation by using smoothness estimates of the solution.

More concretely, the Stein-Chen method can be briefly sketched as follows. An integer-valued random variable Z is Poisson distributed with mean $\lambda > 0$ if and only if

$$E[\lambda g(Z+1) - Zg(Z)] = 0$$

for any bounded function $g : \mathbb{Z}_+ \rightarrow \mathbb{R}$. Thus $E[\lambda g(Z+1) - Zg(Z)] = 0$ is an identity (called *Stein identity*) characterising the Poisson distribution. Let $Z \sim$

$\text{Poi}(\lambda)$ and let W be another integer-valued random variable that we suppose to be almost Poisson distributed. Then we expect that $\mathbb{E}h(W) \approx \mathbb{E}h(Z)$ for a class of test functions h . By the Stein equation whose solution g needs to be determined in terms of h , we have

$$\mathbb{E}h(W) - \mathbb{E}h(Z) = E[\lambda g(Z+1) - Zg(Z)],$$

and it is clear that instead of trying to find a uniform bound on $\mathbb{E}h(W) - \mathbb{E}h(Z)$, we may instead attempt to do so for $E[\lambda g(Z+1) - Zg(Z)]$ by using properties of g . Note that with the choice $h := I_A$, where $A \subseteq \mathbb{Z}_+$, a uniform upper bound on the difference of the expectations $\mathbb{E}h(W) - \mathbb{E}h(Z)$ results in a bound on the total variation distance between the laws of W and Z , i.e. on $d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda)) = \sup_A |P(W \in A) - P(Z \in A)|$.

In standard examples it is assumed that $W = \sum_{i=1}^n I_i$, where the I_i are independent Bernoulli random variables with success probabilities p_i , $i = 1, \dots, n$, or, more simply, i.i.d. Bernoulli random variables with success probability p . Chapter 1 in Barbour et al. (1992) reviews earlier results for bounds on $d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda))$ which were established without the use of Stein's method. One of these is the bound $\sum_{i=1}^n p_i^2$ by Le Cam (1960). It was also obtained by Serfling (1975) by using the simple coupling inequality $d_{TV}(\mathcal{L}(W), \mathcal{L}(Z)) \leq P(W \neq Z)$. Another bound by Le Cam (1960) is $8\lambda^{-1} \sum_{i=1}^n p_i^2$ which was sharpened by Kerstan (1964) to $1.05\lambda^{-1} \sum_{i=1}^n p_i^2$. Compared to these, the Stein-Chen method gives the sharpest bound:

$$(1 - e^{-\lambda})\lambda^{-1} \sum_{i=1}^n p_i^2 \leq \min(1, \lambda^{-1}) \sum_{i=1}^n p_i^2.$$

The main strength of the Stein-Chen method however lies in that it is easily adapted to the case where the Bernoulli variables are no longer independent; see, e.g. Chen (1975b) or Barbour et al. (1992).

By writing the solution g of the Stein equation as the first backward difference of a function γ , i.e. $g(k) = \gamma(k) - \gamma(k-1)$ for all $k \in \mathbb{Z}_+$, Barbour (1988) noticed that the Stein identity for Poisson approximation could be rephrased in terms of the infinitesimal generator of a Markov process with Poisson equilibrium distribution. More precisely,

$$\lambda g(k+1) - kg(k) = \lambda \gamma(k+1) + k\gamma(k-1) - (\lambda + k)\gamma(k) = (\mathcal{A}\gamma)(k),$$

for all $k \in \mathbb{Z}_+$, where \mathcal{A} is the infinitesimal generator of an immigration-death process on \mathbb{Z}_+ with constant immigration rate λ and unit per-capita death

rate. The solution γ to the reformulated Stein equation can then be interpreted in a probabilistic way and coupling arguments may be used to determine smoothness estimates. Barbour and Brown (1992) used this approach to develop the Stein-Chen method for Poisson process approximation. For approximation of the law of a point process Ξ with finite mean measure λ on a locally compact separable metric space E , the Markov process used is a spatial immigration-death process with immigration intensity λ and unit per-capita death rate, whose equilibrium distribution is a Poisson process with mean measure λ (that we denote by $\text{PRM}(\lambda)$). Barbour and Brown (1992) established bounds on $d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\lambda))$ that resemble the results obtained for Poisson approximation, but lack the multiplicative factors depending on $\lambda = \lambda(E)$ which decrease, and thereby reduce the error, as λ increases. The reason for this is that if there are small shifts in the positions of the points of the two processes, the total variation distance takes its maximum value, i.e. 1. An example is given by the following: let I_1, \dots, I_n be independent Bernoulli random variables with success probability $p \in (0, 1)$ and let $W = \sum_{i=1}^n I_i$. Let $\Xi_n = \sum_{i=1}^n I_i \delta_{i/n}$ be a point process on $[0, 1]$ with $\mathbb{E}\Xi_n([0, 1]) = np$. It would be natural to want to compare this process by a Poisson process Ξ with rate np on $[0, 1]$, which can be realised by $\sum_{j=1}^{W^*} \delta_j$, where $W^* \sim \text{Poi}(np)$. However, the Poisson process then takes no points in $n^{-1}\mathbb{Z}_+$ with probability 1, whereas the Bernoulli process Ξ_n takes no points in $n^{-1}\mathbb{Z}_+$ with probability $P(I_1 = 0, \dots, I_n = 0) = (1 - p)^n = \exp\{n \log(1 - p)\} \sim e^{-\lambda}$. Thus, for $R := \{\sum_{j \in n^{-1}\mathbb{Z}_+} \delta_j\}$,

$$d_{TV}(\mathcal{L}(\Xi_n), \mathcal{L}(\Xi)) \geq |P(\Xi_n \notin R) - P(\Xi \notin R)| \sim |1 - e^{-\lambda}| \rightarrow 1,$$

as λ (and n) $\rightarrow \infty$. The total variation distance is thus not suited to approximate a point process with points on a lattice, for instance \mathbb{Z}^d , by a Poisson process with continuous intensity on \mathbb{R}^d , where $d \geq 1$. Barbour and Brown (1992) therefore construct the weaker d_2 -distance which recovers factors in λ that reduce the error. For the special case of marked point processes $\Xi = \sum_{i=1}^n I_i \delta_{X_i}$, smaller errors may yet be obtained in the total variation distance. Indeed, if the marks X_1, \dots, X_n are i.i.d. (and independent of the I_i 's), Ξ can be realised as $\sum_{j=1}^W \delta_{Z_j}$, with Z_j i.i.d., distributed as $\mathcal{L}(X_1)$ and independent of $W = \sum_{i=1}^n I_i$, whereas a Poisson process with mean measure λ can be realised as $\sum_{j=1}^{W^*} \delta_{Z_j}$, with $W^* \sim \text{Poi}(\lambda)$ independent of the Z_j 's. An argument made by Michel (1988) shows that the bounds obtained for Poisson approximation

may then be reused:

$$d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\boldsymbol{\lambda})) \leq d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda)). \quad (1.1)$$

The aim of this thesis is to apply the Stein-Chen method for Poisson and for Poisson process approximation to problems from EVT, with a focus on multivariate extremes. In particular, we are interested in studying random configurations of “extreme points” in space. By “extreme point” we always mean an atypical realisation of a random variable X or a d -dimensional random vector $\mathbf{X} = (X_1, \dots, X_d)$. Throughout the thesis, we suppose that we have a sample of n i.i.d. random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ with state space $E \subseteq \mathbb{R}^d$, where $d \geq 1$, and we denote by $A = A_n$ a measurable subset of E containing the extreme points. For $d = 1$, A is typically of the form (u_n, ∞) , i.e. it contains exceedances of a threshold u_n . For $d > 1$, there is more flexibility as to the choice of “extreme region” A . We might, for instance, set $A := [u_1, \infty) \times \dots \times [u_d, \infty)$ which implies that points in A are extreme in *all* components. We might also define A as the complement of $(-\infty, u_1) \times \dots \times (-\infty, u_d)$; then A not only contains jointly extreme points but also points that might have only one extreme component. We denote the random number of points in A by $W_A = \sum_{i=1}^n I_{\{\mathbf{X}_i \in A\}}$, where $I_{\{\mathbf{X}_i \in A\}}$, $i = 1, \dots, n$, are Bernoulli random variables with probability of success $P(\mathbf{X} \in A)$. Furthermore, we define *marked point processes of exceedances* (MPPE’s) by

$$\Xi_A = \sum_{i=1}^n I_{\{\mathbf{X}_i \in A\}} \delta_{\mathbf{X}_i},$$

where $\delta_{\mathbf{x}}$ denotes Dirac measure on $\mathcal{B}(E)$, and we call $\mathbf{X}_1, \dots, \mathbf{X}_n$ the *marks* of the MPPE. Using (1.1) and results from the Stein-Chen method for Poisson approximation, we establish the following general error estimate:

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \leq P(\mathbf{X} \in A). \quad (1.2)$$

This result serves as the basis for a wide range of applications.

For a first application, let $X_{(n)}$ be the maximum of a sample of n random variables whose distribution function and survival function we denote by F and \bar{F} , respectively. Then (1.2) in particular gives

$$|P(W_A = 0) - \text{Poi}(\mathbb{E}W_A)\{0\}| = \left| P(X_{(n)} \leq u_n) - e^{-n\bar{F}(u_n)} \right| \leq \bar{F}(u_n),$$

which we use to determine the error of the approximation in the Kolmogorov distance of the law of $X_{(n)}$ by that of an extreme value distribution. We achieve this for random variables whose distribution is one of the following: exponential, Pareto, uniform, standard normal, standard Cauchy, geometric. Sometimes the approximation to an extreme value distribution has to be done in more than one step; we then give error bounds for each step. The geometric distribution is a special case here, as it is well known that there is no non-degenerate limit distribution for the distribution of the maximum unless $\lim_{x \uparrow x_F} \bar{F}(x)/\bar{F}(x-) = 1$, where x_F denotes the right endpoint of F . This condition is not satisfied for some of the most well known discrete distributions, in particular, the geometric distribution. However, by letting the parameter(s) of the distribution vary with the sample size n at a suitable rate, it is possible to find a limiting distribution for the law of the maximum. For geometric random variables, we show that we do not actually need the success probability to be varying with n in order to determine a limit law for their maximum, if we allow this limit law to be discrete. More precisely, we approximate the law of the maximum by a discretised Gumbel distribution and determine a bound on the error of this approximation. We also establish an estimate of the error that arises from the approximation by the continuous Gumbel distribution, which will clarify in what way the success probability needs to vary with n for a small error. Furthermore, we determine a better choice of normalising constants than the ones used by Nadarajah and Mitov (2002) for their asymptotic result.

We further apply (1.2) to MPPE's whose marks are distributed according to any of the univariate distributions mentioned above. The error will be given in terms of the chosen threshold, and we discuss the relation between the expected number of threshold exceedances and the size of the error of the approximation in the total variation distance. The bigger the threshold, the smaller the error will be and the fewer exceedances will be expected. We can thus regulate the size of the threshold according to how many of the biggest order statistics we want the MPPE to capture. Moreover, as the aim should be to approximate $\mathcal{L}(\Xi_A)$ by a Poisson process that is easy to use, we determine in each case a practicable intensity function for the approximating Poisson process. Sometimes we would like to approximate $\mathcal{L}(\Xi_A)$ by a Poisson process with a different intensity measure than $\mathbb{E}\Xi_A$, say, $\tilde{\lambda}_A$, which we suppose has a “nicer” intensity function. We then add an error estimate for $d_{TV}(\text{PRM}(\mathbb{E}\Xi_A), \text{PRM}(\tilde{\lambda}_A))$ to the previous error. Again, the geometric distribution is a special case, as the intensity measure of an MPPE with geometric marks is only defined on a lattice. We would prefer to approximate the law of the MPPE with geometric

marks by a Poisson process with continuous intensity function equal to that obtained for the MPPE with exponential marks. As the total variation distance is too strong to achieve this, we instead use the weaker d_2 -distance. For a sharp error estimate, we again need the condition that the success probability of the geometric distribution varies with n at a suitable rate.

We next suppose that $d \geq 2$, i.e. that the marks of the MPPE's are multivariate. We distinguish between sets A that contain only points which exceed thresholds in all d components (which we call “*joint threshold exceedances*”), and sets A that also contain points for which there might be a threshold exceedance in only one component (which we dub “*single-component threshold exceedances*”). It only makes sense to study joint threshold exceedances if the common multivariate distribution of the marks is likely to give rise to joint extremes, though it is not always clear how to find out whether this is the case. One possibility, for $d = 2$, is to compute the coefficient of upper tail dependence. An issue that arises for MPPE's with multivariate marks is thus the choice of the set A . Another issue is that there are infinitely many choices for the dependence structure between the marginal distributions of the marks. Though (1.2) gives a very general result, it might be difficult to understand in what way the error estimate $P(\mathbf{X} \in A)$ behaves and how many threshold exceedances to expect if the joint distribution function has a complicated structure. For both kinds of sets A , we therefore determine easier error bounds for $d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A))$, which are valid for any joint distribution function of the marks. Again, the aim should be to approximate an MPPE by a Poisson process with a practicable intensity function. Also, the Poisson process should, if possible, be independent of the sample size n . In cases where the intensity function of the Poisson process does not meet these requirements, the goal is to approximate by a further Poisson process with a nicer intensity and to determine an error bound on this approximation. Ad hoc considerations are needed to determine whether this is needed, and we therefore have to restrict ourselves to examples. We treat two main examples of MPPE's Ξ_A for which the intensity function of the approximating $\text{PRM}(\mathbb{E}\Xi_A)$ is both too complicated to handle with ease, and varies with n .

First, we study the example of MPPE's whose marks are distributed according to any one out of a subclass of Archimedean copulas that exhibit upper tail dependence. More precisely, we suppose that the generator function ϕ of the Archimedean copulas that we use satisfies $\lim_{r \downarrow 0} r\phi'(1-r)/\phi(1-r) = \theta \in (1, \infty)$. Charpentier and Segers (2009) showed that these copulas have a certain kind of asymptotic tail behaviour, that we make use of to determine a more suitable

intensity function. We proceed by establishing a bound on the error in the total variation distance of the approximation of the Poisson process with mean measure $\mathbb{E}\Xi_A$ by the Poisson process with the new intensity function. Secondly, we consider an MPPE whose marks follow the bivariate Marshall-Olkin geometric distribution. As for univariate geometric marks, the approximating Poisson process lives on a lattice and the aim is to further approximate by a Poisson process with a continuous intensity. We construct a suitable continuous intensity function by spreading the point probabilities of the Marshall-Olkin geometric distribution uniformly over the coordinate squares of the lattice. This intensity function still has to depend on the sample size n if we want the corresponding error to be sharp. We therefore determine its pointwise limit as $n \rightarrow \infty$, under certain conditions on the parameters of the Marshall-Olkin geometric distribution, and use this limit as the new intensity function. We prove error bounds for each step in the d_2 -distance, and, whenever possible, in the total variation distance.

Many more examples could be studied and we hope that this thesis serves as a starting point for more research on the application of the Stein-Chen method for Poisson and Poisson process approximation to topics in EVT. A noticeable limitation of our work is the focus on i.i.d. samples, as one of the strengths of Stein's method is that it is applicable also to random variables that display some dependence on each other; see Section 3.2.3 for a brief discussion.

The structure of the thesis is as follows: **Chapter 2** introduces the Stein-Chen method in detail, both for Poisson approximation and for Poisson process approximation. **Chapter 3** then treats univariate extremes, that is, the approximation of the maximum law of univariate random variables by an extreme value distribution in the Kolmogorov distance, as well as the approximation of MPPE's by suitable Poisson processes in the total variation distance, or, if necessary, the d_2 -distance. **Chapter 4** studies Poisson process approximation for MPPE's with multivariate marks that either have independent components, or components following a certain dependence structure which we describe by way of copulas.

2 The Stein-Chen method

Stein's method provides a way to determine bounds on the errors that arise when approximating one probability distribution by another. This chapter gives an introduction to Stein's method for approximation by a Poisson distribution, as developed by Chen (1975a), as well as for approximation by a Poisson process, as studied by Barbour and Brown (1992). Section 2.1 first discusses the error involved in the law of small numbers, that is, in the approximation of the binomial distribution by the Poisson distribution. Section 2.2 lists some distances between probability measures: the total variation, Kolmogorov, and Wasserstein distances. We later always express the errors of the approximations that we study in one of these distances. Section 2.3 gives an error bound for the approximation, in the total variation distance, of a sum of independent indicator variables by a Poisson distribution. This result by Le Cam (1960) only uses a simple coupling argument. Section 2.4 develops the Stein-Chen method for Poisson approximation in the total variation distance and gives an improvement on the result obtained in Section 2.3. It also treats a result by Chen (1975b) for sums of dependent indicator variables, and outlines two more general procedures proposed by Stein (1986) and Barbour (1988), respectively. Section 2.5 introduces point processes and Poisson processes, and proceeds to develop the Stein-Chen method for approximation by a Poisson process, again in the total variation distance, as achieved by Barbour and Brown (1992). Since errors are worse than for Poisson approximation, Section 2.6 introduces the d_2 -distance by Barbour and Brown (1992), which is weaker than the total variation distance and yields sharper estimates.

2.1 Poisson approximation of the binomial distribution

The Poisson distribution describes the probability of a given number of independent events occurring within a fixed interval of time (or space) when the average rate of occurrence of such events is known from previous observation.

It is named after Siméon Denis Poisson who introduced it in his 1837 treatise *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*, where he showed that the Poisson distribution arises as the limit of a binomial distribution whose probability of success $p = p_n$ varies with the sample size n in such a way that $p_n \rightarrow 0$ and $np_n \rightarrow \lambda > 0$ as $n \rightarrow \infty$. This asymptotic result is called the “law of small numbers” or “law of rare events”. We show that the binomial converges pointwise to the Poisson by considering the probability mass function of a $\text{Bin}(n, p_n)$ -random variable W for any fixed $k \in \mathbb{Z}_+$, and for any integer $n \geq k$:

$$\begin{aligned} P(W = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \prod_{j=0}^{k-1} (n - j) \frac{p_n^k}{k!} e^{(n-k) \log(1-p_n)} \\ &= \prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right) \frac{(np_n)^k}{k!} e^{n \log(1-p_n)} \cdot e^{-k \log(1-p_n)} \quad (2.1) \\ &\sim \frac{(np_n)^k}{k!} e^{-np_n} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{as } n \rightarrow \infty, \end{aligned}$$

since $\log(1 - p_n) \sim -p_n$, as $p_n \rightarrow 0$. In view of this result, it is natural to think of approximating a binomial distribution by a Poisson distribution even for a fixed sample size n , as long as the sample size is large and the success probability is small. But what is the error resulting from such an approximation of a $\text{Bin}(n, p_n)$ distribution by a $\text{Poi}(np_n)$? We may assume that $np_n^2 \leq 1/2$, i.e. $p_n \leq (2n)^{-1/2}$, since else $np_n > (2p_n)^{-1} \rightarrow \infty$ as $p_n \rightarrow 0$. On the one hand, note that

$$\prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right) \leq e^{-\frac{k(k-1)}{2n}},$$

since $1 - j/n \leq e^{-j/n}$ for all $j \geq 0$, and $\sum_{j=0}^{k-1} j/n = k(k-1)/2n$. Moreover,

$$(1 - p_n)^{n-k} \leq e^{-(n-k)\left(p_n + \frac{p_n^2}{2}\right)},$$

since $\log(1 - p_n) \leq -p_n - p_n^2/2$. It follows that $P(W = k)$ from (2.1) is smaller than

$$\frac{(np_n)^k}{k!} e^{-np_n} \cdot \exp \left\{ -\frac{1}{2} \left(\frac{k^2}{n} + np_n^2 \right) + kp_n + \frac{1}{2} \left(\frac{k}{n} + kp_n^2 \right) \right\}.$$

By the inequality of the arithmetic and geometric means, we have

$$\frac{1}{2} \left(\frac{k^2}{n} + np_n^2 \right) \geq \sqrt{\frac{k^2}{n} \cdot np_n^2} = kp_n,$$

and thus

$$P(W = k) - \text{Poi}(np_n)\{k\} \leq \frac{(np_n)^k}{k!} e^{-np_n} \left[\exp \left\{ \frac{1}{2} \left(\frac{k^2}{n} + kp_n^2 \right) \right\} - 1 \right].$$

Since, for $n \geq k$ and $np_n^2 \leq \frac{1}{2}$, we have $\exp \left\{ \frac{1}{2} \left(\frac{k^2}{n} + kp_n^2 \right) \right\} \leq 1 + e^{3/4} \left(\frac{k^2}{n} + np_n^2 \right)$ (using $e^{z/2} \leq 1 + ze^{z/2}$ for $z \geq 0$), it holds that

$$P(W = k) - \text{Poi}(np_n)\{k\} \leq \frac{(np_n)^k}{k!} e^{-np_n} \cdot e^{3/4} \left(\frac{k^2}{n} + np_n^2 \right). \quad (2.2)$$

On the other hand, since $np_n^2 \leq 1/2$, we have $p_n \leq 1/2$ for all $n \geq 2$, and therefore $(1 - p_n)^{-1} \leq 1 + 2p_n$. It follows that

$$e^{n \log(1-p_n)} \geq e^{-\frac{np_n}{1-p_n}} \geq e^{-np_n - 2np_n^2} \geq e^{-np_n} (1 - 2np_n^2).$$

Moreover, an induction proof readily shows that $\prod_{j=0}^{k-1} (1 - j/n) \geq 1 - k^2/n$. Thus,

$$\begin{aligned} \text{Poi}(np_n)\{k\} - P(W = k) &\leq \frac{(np_n)^k}{k!} e^{-np_n} \left\{ 1 - (1 - 2np_n^2) \left(1 - \frac{k^2}{n} \right) \right\} \\ &\leq \frac{(np_n)^k}{k!} e^{-np_n} \left(\frac{k^2}{n} + 2np_n^2 \right). \end{aligned} \quad (2.3)$$

It follows from (2.2) and (2.3) that, for $k \leq n$ and $np_n^2 \leq 1/2$,

$$|P(W = k) - \text{Poi}(np_n)\{k\}| \leq e^{3/4} \frac{(np_n)^k}{k!} e^{-np_n} \left(\frac{k^2}{n} + np_n^2 \right). \quad (2.4)$$

Note that we have not used $np_n \rightarrow \lambda$, as $n \rightarrow \infty$, in order to establish (2.4), and that (2.4) is stronger than just a limit result: it gives an explicit estimate of the error of the approximation of a binomial by a Poisson point probability at k for any sample size $n \geq k$. By (2.4), the approximation is good so long as the terms np_n^2 and $k^2 n^{-1}$ are small. We can aim for an even stronger

statement than (2.4) by investigating the accuracy of the approximation of one probability distribution by another in a probability metric, such as the total variation distance that we will define below. For instance, for any subset $A \subseteq \{0, 1, \dots, n\}$, (2.4) gives

$$\begin{aligned}
 & |P(W \in A) - \text{Poi}(np_n)\{A\}| \\
 & \leq e^{3/4} e^{-np_n} \sum_{k \in A} \frac{(np_n)^k}{k!} \left(\frac{k^2}{n} + np_n^2 \right) \\
 & \leq e^{3/4} e^{-np_n} \left\{ np_n^2 \sum_{k=0}^{\infty} \frac{(np_n)^k}{k!} + p_n \sum_{k=1}^{\infty} \frac{(np_n)^{k-1}}{(k-1)!} k \right\} \\
 & = e^{3/4} \{2np_n^2 + p_n\}, \tag{2.5}
 \end{aligned}$$

where we used

$$\sum_{k=1}^{\infty} \frac{(np_n)^{k-1}}{(k-1)!} k = \sum_{l=0}^{\infty} \frac{(np_n)^l}{l!} (l+1) = np_n \sum_{l=1}^{\infty} \frac{(np_n)^{l-1}}{(l-1)!} + \sum_{l=0}^{\infty} \frac{(np_n)^l}{l!}$$

and $\sum_{k=0}^{\infty} (np_n)^k / k! = e^{np_n}$. Furthermore, note that $P(W \in [n+1, \infty)) = 0$, and that for all $n \geq 2$, $\text{Poi}(np_n)\{[n+1, \infty)\}$ is of smaller order than the bound in (2.5). More precisely, Proposition A.2.3 in Barbour et al. (1992), gives, for all $n \geq 2$,

$$\begin{aligned}
 \text{Poi}(np_n)\{[n+1, \infty)\} & \leq \frac{(n+2)e^{-\frac{(n+1-np_n)^2}{2(n+1+np_n)}}}{(n+2-np_n)\sqrt{2\pi(n+1)}} \leq \frac{e^{-\frac{n}{2}\left(\frac{1}{2}+\frac{1}{n}\right)^2/\left(\frac{3}{2}+\frac{1}{n}\right)}}{\left(1-\frac{np_n}{n+2}\right)\sqrt{2\pi n}} \\
 & \leq \left(1+\frac{n}{n+2}\right) \frac{e^{-\frac{n}{4}\left(\frac{1}{4}+\frac{1}{n}+\frac{1}{n^2}\right)}}{\sqrt{2\pi n}} \leq \frac{e^{-\frac{n}{16}}}{\sqrt{n}},
 \end{aligned}$$

where we also used that $p_n \leq 1/2$ for $n \geq 2$, and $(1-z)^{-1} \leq 1+2z$ for $z = np_n/(n+2)$. For any $A \subseteq \mathbb{Z}_+$, the error $|P(W \in A) - \text{Poi}(np_n)\{A\}|$ (and thereby the total variation distance between $\mathcal{L}(W)$ and $\text{Poi}(np_n)$; see Section 2.2 below) is then at most of order $\max(np_n^2, p_n)$. The approximation between the $\text{Bin}(n, p_n)$ and the $\text{Poi}(np_n)$ is therefore sharp if $p_n = o(n^{-1/2})$.

2.2 Distances between probability measures

In general the aim is to find an upper bound for the difference between the expectations of a test function from a predetermined family of test functions under the two distributions. Each family of test functions determines an associated metric. We list three important examples of such distances between probability measures: the total variation, the Kolmogorov and the Wasserstein distances. For each of these, suppose that μ and ν are two probability measures on a measurable space (E, \mathcal{E}) .

Total variation distance

The *total variation distance* between μ and ν is defined as follows:

$$d_{TV}(\mu, \nu) = \sup_{h \in H} \left| \int_E h d\mu - \int_E h d\nu \right| = \sup_{B \in \mathcal{E}} |\mu(B) - \nu(B)|,$$

where the test functions h are indicators of measurable subsets of E , i.e. $H := \{I_B; B \in \mathcal{E}\}$, where, for any $x \in E$, $I_B(x) = 1$ if $x \in B$ and $I_B(x) = 0$ if $x \notin B$. Note that for any set $B \in \mathcal{E}$, we have $\mu(B) - \nu(B) = \nu(B^C) - \mu(B^C)$, with $B^C \in \mathcal{E}$ the complement of B . Therefore,

$$d_{TV}(\mu, \nu) = \sup_{B \in \mathcal{E}} \{\mu(B) - \nu(B)\}. \quad (2.6)$$

An equivalent definition of the total variation distance is given by

$$d_{TV}(\mu, \nu) = \sup_{\tilde{h} \in \tilde{\mathcal{H}}} \left| \int_E \tilde{h} d\mu - \int_E \tilde{h} d\nu \right|, \quad (2.7)$$

where $\tilde{\mathcal{H}} := \{\tilde{h} : E \rightarrow \mathbb{R}; 0 \leq h(x) \leq 1, \text{ for all } x \in E\}$. In order to see that the two definitions are equivalent, note first that any $\tilde{h} \in \tilde{\mathcal{H}}$ may be defined as $\tilde{h}(x) := h(x) - q$, for any $x \in E$, $h \in \mathcal{H}$, and for any choice of $q \in [0, 1]$. Then,

$$\begin{aligned} \sup_{\tilde{h} \in \tilde{\mathcal{H}}} \left| \int_E \tilde{h} d\mu - \int_E \tilde{h} d\nu \right| &= \sup_{h \in H} \left| \int_E (h - q) d\mu - \int_E (h - q) d\nu \right| \\ &= \sup_{h \in H} \left| \int_E h d\mu - q\mu(E) - \int_E h d\nu + q\nu(E) \right| \\ &= \sup_{h \in H} \left| \int_E h d\mu - \int_E h d\nu \right|, \end{aligned}$$

since $\mu(E) = \nu(E) = 1$. Further equivalent definitions can be found in Barbour et al. (1992), pp. 253-254. Note also that if E is a separable metric space and if X and Y are two E -valued random variables defined on the same probability space with distributions μ and ν , respectively, then, for $B \in \mathcal{E}$,

$$\begin{aligned}\mu(B) - \nu(B) &= P(X \in B) - P(Y \in B) = \mathbb{E}[I_B(X) - I_B(Y)] \\ &\leq \mathbb{E}I_{\{X \neq Y\}} = P(X \neq Y),\end{aligned}$$

and therefore

$$d_{TV}(\mu, \nu) = \sup_{B \in \mathcal{E}} |P(X \in B) - P(Y \in B)| \leq P(X \neq Y). \quad (2.8)$$

Remark 2.2.1. We used a *coupling* of the two distributions μ and ν in order to establish (2.8). Throughout this work, we will time and again use couplings. In general, coupling means the joint construction of two random variables (or processes) X and Y which (marginally) follow two distributions μ and ν of interest. Its purpose is to relate the two previously unrelated distributions μ and ν in some way so as to be able to compare them. For details on the coupling method, consult Lindvall (2002) or Thorisson (2000).

Remark 2.2.2. Suppose that $(X_n)_{n \in \mathbb{N}}$ is a sequence of random variables with a discrete state space E , for instance $E = \mathbb{Z}_+$. This sequence then converges in total variation to a random variable X , i.e. $d_{TV}(\mathcal{L}(X_n), \mathcal{L}(X)) \rightarrow 0$ as $n \rightarrow \infty$, if and only if $P(X_n = k) \rightarrow P(X = k)$ for all $k \in E$ as $n \rightarrow \infty$, that is, if and only if it converges in distribution (or weakly) to X . We can see this by noting that, on the one hand, convergence in distribution follows from convergence in total variation, since, for any $k \in E$,

$$\begin{aligned}0 &\leq |P(X_n = k) - P(X = k)| \\ &\leq \sup_{B \subseteq E} |P(X_n \in B) - P(X \in B)| \\ &= d_{TV}(\mathcal{L}(X_n), \mathcal{L}(X)).\end{aligned}$$

On the other hand, suppose that $P(X_n = k) \rightarrow P(X = k)$ for all $k \in E$, as $n \rightarrow \infty$, define $\tilde{B} := \{k \in E : P(X_n = k) \geq P(X = k)\}$, and note that

$d_{TV}(\mathcal{L}(X_n), \mathcal{L}(X))$ equals

$$\begin{aligned}
 & P(X_n \in \tilde{B}) - P(X \in \tilde{B}) \\
 &= \frac{1}{2} \left\{ P(X_n \in \tilde{B}) - P(X \in \tilde{B}) + P(X \notin \tilde{B}) - P(X_n \notin \tilde{B}) \right\} \\
 &= \frac{1}{2} \left\{ \sum_{k \in \tilde{B}} [P(X_n = k) - P(X = k)] + \sum_{k \in \tilde{B}^C} [P(X = k) - P(X_n = k)] \right\} \\
 &= \frac{1}{2} \sum_{k \in E} |P(X_n = k) - P(X = k)|.
 \end{aligned}$$

For each $\epsilon > 0$, there exists a finite set $K \subseteq E$ such that $\sum_{k \in K^C} P(X = k) \leq \epsilon/4$, and there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $\sum_{k \in K} |P(X_n = k) - P(X = k)| \leq \epsilon/4$. Convergence in total variation then follows from convergence in distribution, since, for all $n \geq n_0$,

$$\begin{aligned}
 & \sum_{k \in E} |P(X_n = k) - P(X = k)| \\
 &= \sum_{k \in K} |P(X_n = k) - P(X = k)| + \sum_{k \in K^C} |P(X_n = k) - P(X = k)| \\
 &\leq \frac{\epsilon}{4} + \sum_{k \in K^C} P(X_n = k) + \sum_{k \in K^C} P(X = k) = \frac{\epsilon}{4} + 1 - \sum_{k \in K} P(X_n = k) + \frac{\epsilon}{4} \\
 &\leq \frac{\epsilon}{4} + 1 - \sum_{k \in K} P(X = k) + \sum_{k \in K} |P(X = k) - P(X_n = k)| + \frac{\epsilon}{4} \\
 &\leq \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon.
 \end{aligned}$$

If E is not discrete, for example, if $E = \mathbb{R}$, then convergence in total variation is stronger than convergence in distribution, and it might occasionally be even too strong to be of use.

Kolmogorov distance

Suppose that $E = \mathbb{R}$ and that the test functions are the indicators of half-lines in \mathbb{R} , i.e. $\mathcal{H} = \{I_{(-\infty, x]}; x \in \mathbb{R}\}$. The *Kolmogorov distance* is then defined as follows:

$$d_K(\mu, \nu) = \sup_{h \in \mathcal{H}} \left| \int_E h d\mu - \int_E h d\nu \right| = \sup_{x \in \mathbb{R}} |\mu\{(-\infty, x]\} - \nu\{(-\infty, x]\}|.$$

For two random variables $X \sim \mu$ and $Y \sim \nu$, the Kolmogorov distance is thus given by $\sup_{x \in \mathbb{R}} |P(X \leq x) - P(Y \leq x)|$.

Wasserstein distance induced by d

Suppose that E is a separable metric space with associated metric d and equipped with its Borel σ -field. For the *Wasserstein distance*, which is also known as the *Dudley, Fortet-Mourier* or *Kantorovich distance*, we only consider probability measures μ such that for some, and then for any, $x_0 \in E$, $\mathbb{E}d(X, x_0) = \int_E d(x, x_0) d\mu(x) < \infty$, where $X \sim \mu$. For probability measures μ and ν satisfying this condition, the Wasserstein distance induced by d is given by

$$d_W(\mu, \nu) = \sup_{h \in \mathcal{H}} \left| \int_E h d\mu - \int_E h d\nu \right|,$$

where the test functions h are uniformly Lipschitz with constant 1, i.e. $\mathcal{H} = \{h : E \rightarrow \mathbb{R}, |h(x) - h(y)| \leq d(x, y)\}$. It can be shown that

$$d_W(\mu, \nu) = \inf \mathbb{E}d(X, Y), \quad (2.9)$$

where the infimum is taken over all couplings (X, Y) of μ and ν (see, for instance, Section 7.1 in Ambrosio et al. (2005)).

Chapter 2 mainly uses the total variation distance. The Wasserstein distance will make its first appearance only in Section 2.6 where we will use it to construct a distance between probability measures over a set of point measures, in the context of approximation by Poisson processes. The Kolmogorov distance will be widely used in Section 3.1 where we approximate the law of the maximum of i.i.d. random variables by an extreme value distribution.

2.3 Le Cam's result for Poisson approximation

Le Cam (1960) determined an upper bound for the accuracy in total variation of the approximation of a sum of independent Bernoulli random variables by a Poisson distribution with the same mean. We give the argument by Serfling (1975) that uses a simple coupling inequality.

Theorem 2.3.1. (*Le Cam, 1960*) Let I_1, \dots, I_n be independent Bernoulli random variables with $P(I_i = 1) = p_i$ and $P(I_i = 0) = 1 - p_i$, where $0 < p_i < 1$,

for all $i = 1, \dots, n$. Let $W := \sum_{i=1}^n I_i$ and $\lambda := \mathbb{E}W = \sum_{i=1}^n p_i$. Then

$$d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda)) \leq \sum_{i=1}^n p_i^2.$$

Proof. We perform a coupling of the two distributions $\mathcal{L}(W)$ and $\text{Poi}(\lambda)$ by defining random variables J_i and Y_i on probability spaces (Ω_i, P_i) , for each $i = 1, \dots, n$, such that $\sum_{i=1}^n J_i$ and $\sum_{i=1}^n Y_i$ follow the distributions $\mathcal{L}(W)$ and $\text{Poi}(\lambda)$, respectively. To achieve this, define

$$\begin{aligned} \Omega_i &:= \{-1, 0, 1, 2, \dots\}, \\ P_i(0) &:= 1 - p_i, \\ P_i(k) &:= e^{-p_i} p_i^k / k!, \text{ for all } k \geq 1, \\ P_i(-1) &:= 1 - P_i(0) - \sum_{k \geq 1} P_i(k) = e^{-p_i} - (1 - p_i), \end{aligned}$$

for all $i = 1, \dots, n$. By construction, (Ω_i, P_i) are probability spaces. Let (Ω, P) be the product space of the probability spaces (Ω_i, P_i) , i.e. let $\Omega := \Omega_1 \times \dots \times \Omega_n$ and, for any $\omega = (\omega_1, \dots, \omega_n) \in \Omega$, define $P(\omega) := P_1(\omega_1) \cdot \dots \cdot P_n(\omega_n)$. Then $\sum_{\omega \in \Omega} P(\omega) = 1$. Now define, for any $\omega \in \Omega$,

$$J_i(\omega) := \begin{cases} 0, & \text{if } \omega_i = 0, \\ 1, & \text{else,} \end{cases}$$

and

$$Y_i(\omega) := \begin{cases} k, & \text{if } \omega_i = k, k \geq 1, \\ 0, & \text{else.} \end{cases}$$

By definition, the random variables J_i have the same distribution as the Bernoulli random variables I_i , i.e. $P(J_i = 1) = p_i = 1 - P(J_i = 0)$, for all $i = 1, \dots, n$, and, by definition of the product space, they are independent. Moreover, the random variables Y_i are independent and follow the $\text{Poi}(p_i)$ -distribution, since $P(Y_i = k) = P_i(k)$ for all $k \geq 1$, and $P(Y_i = 0) = P_i(-1) + P_i(0) = e^{-p_i}$. It follows that the random variable $Y := Y_1 + \dots + Y_n$ is $\text{Poi}(\lambda)$ -distributed. The random variables J_i and Y_i now take the same values if $\omega_i \in \{0, 1\}$ and thus, $P(I_i = Y_i) = P_i(0) + P_i(1) = (1 - p_i) + p_i e^{-p_i}$, and

$$P(I_i \neq Y_i) = p_i (1 - e^{-p_i}) \leq p_i^2,$$

for all $i = 1, \dots, n$, since $1 - e^{-z} \leq z$ for $z > 0$. From (2.8), it now follows that

$$d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda)) \leq P(W \neq Y) \leq \sum_{i=1}^n P(I_i \neq Y_i) \leq \sum_{i=1}^n p_i^2.$$

□

For the accuracy in total variation between the binomial distribution and a Poisson distribution with the same mean, it then follows immediately:

Corollary 2.3.2.

$$d_{TV}(\text{Bin}(n, p), \text{Poi}(np)) \leq np^2.$$

For a good approximation of a binomial distribution by a Poisson, we thus need the probability of success p to vary with the sample size n such that $np^2 \rightarrow 0$ as $n \rightarrow \infty$, i.e. we need $p = p_n = o(n^{-1/2})$ as $n \rightarrow \infty$. We already noticed this in Section 2.1. The error np^2 can, however, still be improved. In order to get an inkling of why this is the case, note first that for large np , most realisations of the $\text{Bin}(n, p)$ and $\text{Poi}(np)$ distributions lie in the vicinity of np . With the following inequalities that can be established for $n \geq 2$ and $p = p_n \leq 1/2$, and that are (in part) more precise than the ones used in Section 2.1,

$$\begin{aligned} e^{-\frac{k^2}{2n} + \frac{k}{2n} - \frac{2k^3}{3n^2}} &\leq \prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right) \leq e^{-\frac{k^2}{2n} + \frac{k}{2n}}, \\ e^{-np} \cdot e^{-\frac{np^2}{2} - 8np^3} &\leq (1-p)^n \leq e^{-np} \cdot e^{-\frac{np^2}{2}}, \\ e^{kp} &\leq (1-p)^{-k} \leq e^{kp + \frac{kp^2}{2}}, \end{aligned}$$

it turns out that for $k = np$,

$$P(W = k) = \{1 + O(p, np^3)\} \text{Poi}(np)\{k\},$$

which suggests that the error estimate from Corollary 2.3.2 may be reduced. With Stein's method we indeed find an improved result; see Corollary 2.4.4 in the next section.

2.4 The Stein-Chen method for Poisson approximation

Stein's method for Poisson approximation was first worked out by Chen (1975a) and is therefore usually named the Stein-Chen method. We demonstrate the method for the example of sums of independent, but non-identically distributed indicator variables, which will provide us with an improvement on the result by Le Cam in Theorem 2.3.1. First note the following two observations, summarised in Theorems 2.4.1 and 2.4.2:

Theorem 2.4.1. (Chen, 1975a) *Let $f : \mathbb{Z}_+ \rightarrow \mathbb{R}$ be a bounded function. The following are equivalent:*

(i) *There exists a bounded solution $g = g_{f,\lambda} : \mathbb{Z}_+ \rightarrow \mathbb{R}$ of*

$$\lambda g(k+1) - kg(k) = f(k), \quad \text{for all } k \in \mathbb{Z}_+. \quad (2.10)$$

(ii) $\mathbb{E}f(Z) = 0$, for $Z \sim \text{Poi}(\lambda)$.

Furthermore, if either is satisfied, then the solution $g = g_{f,\lambda}$ to equation (2.10) is

$$g(k+1) = \frac{k!}{\lambda^{k+1}} \sum_{j=0}^k p_{j,\lambda} e^\lambda f(j) = -\frac{k!}{\lambda^{k+1}} \sum_{j=k+1}^{\infty} p_{j,\lambda} e^\lambda f(j), \quad (2.11)$$

for all $k \in \mathbb{Z}_+$, where $p_{j,\lambda} = \frac{\lambda^j}{j!} e^{-\lambda}$.

Proof. (i) \Rightarrow (ii): Let $g : \mathbb{Z}_+ \rightarrow \mathbb{R}$ be a bounded function and let $Z \sim \text{Poi}(\lambda)$, with $\lambda > 0$. Then

$$\mathbb{E}\lambda g(Z+1) = \sum_{k \geq 0} \lambda g(k+1) \frac{\lambda^k}{k!} e^{-\lambda},$$

and

$$\begin{aligned} \mathbb{E}Zg(Z) &= \sum_{l \geq 1} lg(l) \frac{\lambda^l}{l!} e^{-\lambda} \\ &= \sum_{k \geq 0} \lambda(k+1)g(k+1) \frac{\lambda^k}{(k+1)!} e^{-\lambda} \\ &= \sum_{k \geq 0} \lambda g(k+1) \frac{\lambda^k}{k!} e^{-\lambda}. \end{aligned}$$

It follows from (2.10) that

$$\mathbb{E}f(Z) = \mathbb{E}[\lambda g(Z+1) - Zg(Z)] = 0.$$

Note that we have not needed g to be of the form (2.11) here. (ii) \Rightarrow (i): Suppose $f : \mathbb{Z}_+ \rightarrow \mathbb{R}$ is a bounded function such that $\mathbb{E}f(Z) = 0$ for $Z \sim \text{Poi}(\lambda)$. We use an induction proof to show that $g = g_{f,\lambda}$ given by (2.11) solves equation (2.10). Without loss of generality we set $g(0) := 0$. Then, for $k = 0$, (2.11) gives $g(1) = f(0)p_{0,\lambda}e^\lambda/\lambda = f(0)/\lambda$, which solves (2.10). Assume that (2.11) solves (2.10) for an integer $k \geq 0$. Then the induction step

$$\begin{aligned} g(k+2) &= \frac{(k+1)!}{\lambda^{k+2}} \sum_{j=0}^{k+1} p_{j,\lambda} e^\lambda f(j) \\ &= \frac{k+1}{\lambda} \cdot \frac{k!}{\lambda^{k+1}} \left\{ \sum_{j=0}^k p_{j,\lambda} e^\lambda f(j) + p_{k+1,\lambda} e^\lambda f(k+1) \right\} \\ &= \lambda^{-1} \{ (k+1)g(k+1) + f(k+1) \} \end{aligned}$$

completes the argument. Note that we have not yet needed the condition $\mathbb{E}f(Z) = 0$, which implies that the solution g we found so far always exists. The condition is needed, however, for the alternative representation of $g(k+1)$ as will be made clear by the following:

$$0 = \mathbb{E}f(Z) = \sum_{j=0}^k \frac{\lambda^j}{j!} e^{-\lambda} f(j) + \sum_{j=k+1}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} f(j).$$

Multiplication of both sides by $k!e^\lambda/\lambda^{k+1}$ now yields

$$\frac{k!}{\lambda^{k+1}} \sum_{j=0}^k p_{j,\lambda} e^\lambda f(j) = -\frac{k!}{\lambda^{k+1}} \sum_{j=k+1}^{\infty} p_{j,\lambda} e^\lambda f(j).$$

Furthermore, it follows from (2.11) that

$$\begin{aligned} |g(k+1)| &\leq \|f\| \sum_{j=k+1}^{\infty} \frac{k! \lambda^{j-(k+1)}}{j!} \\ &= \|f\| \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(k+m) \cdot \dots \cdot (k+1)} \\ &\leq \|f\| \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{m!}, \end{aligned}$$

since, obviously, $k+m \geq m$ for all $m \geq 1$. Moreover,

$$\sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{m!} \leq \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} = e^{\lambda},$$

and thus $\|g\| \leq e^{\lambda} \|f\|$, i.e. g is bounded. \square

We can in fact give a characterisation of the Poisson distribution:

Theorem 2.4.2. (*Characterisation of the Poisson distribution*) *Let Z be a random variable taking values in \mathbb{Z}_+ . The following are equivalent:*

- (i) $Z \sim \text{Poi}(\lambda)$.
- (ii) *For every bounded function $g : \mathbb{Z}_+ \rightarrow \mathbb{R}$, we have*

$$\mathbb{E}[\lambda g(Z+1) - Zg(Z)] = 0. \quad (2.12)$$

Proof. (i) \Rightarrow (ii): Suppose that $Z \sim \text{Poi}(\lambda)$. Let $h : \mathbb{Z}_+ \rightarrow \mathbb{R}$ be any bounded function and define

$$f(k) := h(k) - e^{-\lambda} \sum_{k=0}^{\infty} h(k) \frac{\lambda^k}{k!}, \quad \text{for all } k \in \mathbb{Z}_+.$$

Then f is bounded and $\mathbb{E}f(Z) = 0$. By Theorem 2.4.1 there exists a bounded function $g = g_{f,\lambda} : \mathbb{Z}_+ \rightarrow \mathbb{R}$ satisfying (2.10), and thus,

$$\mathbb{E}[\lambda g(Z+1) - Zg(Z)] = \mathbb{E}f(Z) = 0.$$

(ii) \Rightarrow (i): Suppose that Z is a random variable with state space \mathbb{Z}_+ satisfying (2.12). Let $X \sim \text{Poi}(\lambda)$, choose any subset $A \subseteq \mathbb{Z}_+$, and define the bounded function $h(k) := h_A(k) := I_{\{k \in A\}}$ for all $k \in \mathbb{Z}_+$. Define

$$f(k) := f_A(k) := h_A(k) - \mathbb{E}h_A(X), \quad \text{for all } k \in \mathbb{Z}_+.$$

Then f is bounded, $\mathbb{E}f(X) = 0$, and by Theorem 2.4.1 there exists a bounded function $g := g_{f,\lambda} : \mathbb{Z}_+ \rightarrow \mathbb{R}$ satisfying (2.10) for all $k \in \mathbb{Z}_+$. By (2.12), we thus obtain

$$\begin{aligned} 0 &= \mathbb{E}[\lambda g(Z+1) - Zg(Z)] = \mathbb{E}f(Z) \\ &= \mathbb{E}h_A(Z) - \mathbb{E}h_A(X) = P(Z \in A) - P(X \in A), \end{aligned}$$

and therefore $Z \stackrel{d}{=} X$, i.e. $Z \sim \text{Poi}(\lambda)$. □

In Section 2.4.1 below, we construct the Stein equation for Poisson approximation, by using Theorem 2.4.1, and give smoothness estimates of its solution. Section 2.4.2 then applies the results to the problem of determining a bound on the error that arises with the approximation of the law of a sum of independent Bernoulli random variables by a Poisson distribution. Section 2.4.3 does the same for sums of dependent Bernoulli variables. Section 2.4.4 introduces a general procedure that uses an exchangeable pair of random variables and an antisymmetric function for establishing a Stein equation. Section 2.4.5 relates the Stein equation for Poisson approximation to generators of immigration-death processes, which allows for a probabilistic interpretation of its solution.

2.4.1 Construction of the Stein equation and smoothness estimates

Let $Z \sim \text{Poi}(\lambda)$ and let $A \subseteq \mathbb{Z}_+$. Define

$$f(k) := f_A(k) := I_{\{k \in A\}} - \text{Poi}(\lambda)\{A\}, \quad k \in \mathbb{Z}_+,$$

where $\text{Poi}(\lambda)\{A\} = P(Z \in A)$. The function f_A is obviously bounded by 1 and we have

$$\mathbb{E}f_A(Z) = P(Z \in A) - \text{Poi}(\lambda)\{A\} = 0.$$

Thus, by Theorem 2.4.1, there exists a bounded solution $g := g_{f_A,\lambda} := g_{A,\lambda} : \mathbb{Z}_+ \rightarrow \mathbb{R}$ to the *Stein equation*

$$\lambda g(k+1) - kg(k) = I_{\{k \in A\}} - \text{Poi}(\lambda)\{A\}, \quad k \in \mathbb{Z}_+, \quad (2.13)$$

and the *Stein solution* is given by plugging $f_A(k)$ into (2.11):

$$\begin{aligned} g(k+1) &= \frac{k!}{\lambda^{k+1}} e^\lambda \{P(Z \in A, Z \leq k) - P(Z \in A)P(Z \leq k)\} \\ &= \frac{k!}{\lambda^{k+1}} e^\lambda \{\text{Poi}(\lambda)\{A \cap U_k\} - \text{Poi}(\lambda)\{A\}\text{Poi}(\lambda)\{U_k\}\}, \end{aligned}$$

where $U_k := \{0, 1, \dots, k\}$, for all $k \in \mathbb{Z}_+$. Let W be a random variable taking values in \mathbb{Z}_+ . By the Stein equation (2.13), taking expectations, we have

$$\mathbb{E}[\lambda g(W+1) - Wg(W)] = P(W \in A) - \text{Poi}(\lambda)\{A\}. \quad (2.14)$$

In order to find an upper bound for the error in total variation of the approximation of the law of W by that of a Poisson distribution with mean $\lambda > 0$, it suffices, by (2.14), to bound $\mathbb{E}[\lambda g(W+1) - Wg(W)]$ uniformly in $A \subseteq \mathbb{Z}_+$. To achieve this, we first need smoothness estimates of the function $g = g_{A,\lambda}$ as given in Lemma 2.4.3. For the proofs of (i) and (ii) of Lemma 2.4.3, we refer to Barbour et al. (1992) (Remark 10.2.4) and Barbour and Eagleson (1983), respectively.

Lemma 2.4.3. *For the solution $g := g_{A,\lambda}$ of the Stein equation, it holds that*

$$(i) \quad \|g\| := \sup_{k \geq 0} |g(k)| \leq \min \left(1, \sqrt{\frac{2}{e\lambda}} \right),$$

$$(ii) \quad \Delta g := \sup_{k \geq 0} |g(k+1) - g(k)| \leq \frac{1 - e^{-\lambda}}{\lambda} \leq \min \left(1, \frac{1}{\lambda} \right).$$

□

2.4.2 Independent indicator variables

We use (2.14) to find a sharper bound than the one given by Le Cam in Theorem 2.3.1. We assume the setting of Theorem 2.3.1, i.e. let I_1, \dots, I_n be independent Bernoulli random variables with $P(I_i = 1) = p_i$ and $P(I_i = 0) = 1 - p_i$, where $0 < p_i < 1$, for all $i = 1, \dots, n$. Let $W = \sum_{i=1}^n I_i$ and $\lambda = \mathbb{E}W = \sum_{i=1}^n p_i$. Then

$$|P(W \in A) - \text{Poi}(\lambda)\{A\}| \leq 2\|g\| \sum_{i=1}^n p_i^2, \quad (2.15)$$

$$|P(W \in A) - \text{Poi}(\lambda)\{A\}| \leq \Delta g \sum_{i=1}^n p_i^2, \quad (2.16)$$

where $\|g\| := \sup_{k \geq 0} |g(k)|$ and $\Delta g := \sup_{k \geq 0} |g(k+1) - g(k)|$. We may show (2.15) and (2.16) as follows: for each $i = 1, \dots, n$, define $W_i := \sum_{j=1, j \neq i}^n I_j$. We have

$$\mathbb{E}[I_i g(W)] = \mathbb{E}[I_i g(W_i + 1)] = p_i \mathbb{E}[g(W_i + 1)], \quad (2.17)$$

since I_i and W_i are independent. We may thus write the left hand side of (2.14) as

$$\begin{aligned}\mathbb{E}[\lambda g(W+1) - Wg(W)] &= \sum_{i=1}^n p_i \{ \mathbb{E}[g(W+1)] - \mathbb{E}[g(W_i+1)] \} \\ &= \sum_{i=1}^n p_i \sum_{k \geq 0} g(k+1) \{ P(W=k) - P(W_i=k) \},\end{aligned}$$

where, using independence between W_i and I_i , we find that

$$\begin{aligned}P(W=k) &= P(W_i + I_i = k, I_i = 0) + P(W_i + I_i = k, I_i = 1) \\ &= (1 - p_i)P(W_i = k) + p_i P(W_i + 1 = k).\end{aligned}$$

We thus obtain

$$\mathbb{E}[\lambda g(W+1) - Wg(W)] = \sum_{i=1}^n p_i^2 \sum_{k \geq 0} g(k+1) \{ P(W_i+1=k) - P(W_i=k) \}. \quad (2.18)$$

On the one hand, for (2.15), this may be bounded from above by

$$\|g\| \sum_{i=1}^n p_i^2 \left\{ \sum_{k \geq 0} P(W_i+1=k) + \sum_{k \geq 0} P(W_i=k) \right\} \leq 2\|g\| \sum_{i=1}^n p_i^2.$$

On the other hand, for (2.16), note that

$$\begin{aligned}\sum_{k \geq 0} g(k+1)P(W_i = k-1) &= \sum_{k \geq 1} g(k+1)P(W_i = k-1) \\ &= \sum_{k' \geq 0} g(k'+2)P(W_i = k').\end{aligned}$$

Then, (2.18) gives

$$\begin{aligned}\left| \sum_{i=1}^n p_i^2 \sum_{k \geq 0} P(W_i = k) \{g(k+2) - g(k+1)\} \right| &\leq \Delta g \sum_{i=1}^n p_i^2 \sum_{k \geq 0} P(W_i = k) \\ &\leq \Delta g \sum_{i=1}^n p_i^2.\end{aligned}$$

The results now follow using (2.14). With (2.14), (2.16) and Lemma 2.4.3, we get an improvement on Le Cam's result and thereby also better rates for the approximation of a binomial by a Poisson distribution:

Theorem 2.4.4. (*Barbour and Hall, 1984*) Let I_1, \dots, I_n be independent Bernoulli random variables with $P(I_i = 1) = p_i$ and $P(I_i = 0) = 1 - p_i$, where $0 < p_i < 1$, for all $i = 1, \dots, n$. Let $W = \sum_{i=1}^n I_i$ and $\lambda = \mathbb{E}W = \sum_{i=1}^n p_i$. Then

$$d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i=1}^n p_i^2 \leq \min\left(1, \frac{1}{\lambda}\right) \sum_{i=1}^n p_i^2. \quad (2.19)$$

□

Corollary 2.4.5. $d_{TV}(\text{Bin}(n, p), \text{Poi}(np)) \leq np^2 \min\left(1, \frac{1}{np}\right) = O(p)$.

2.4.3 Dependent indicator variables – The local approach

Theorems 2.3.1 and 2.4.4 generalise the problem of the approximation of a binomial distribution by a Poisson in the sense that the indicator variables need no longer be identically distributed. One of the strengths of the Stein-Chen method is that we can further relax the conditions put on the indicator variables by dropping the assumption of independence. Indeed, independence is used only once, in Equation (2.17) and thus it is only (2.17) that needs to be modified in a way to allow for some kind of dependence. One of the ways to do this was suggested by Chen (1975b). For each of the indicator variables I_i , $i = 1, \dots, n$, the idea is to classify the $n - 1$ remaining indicator variables into two different categories, those “strongly” dependent on I_i and those “weakly” dependent on I_i .

Theorem 2.4.6. Let I_1, \dots, I_n be Bernoulli random variables with $P(I_i = 1) = p_i$ and $P(I_i = 0) = 1 - p_i$, where $0 < p_i < 1$, for all $i = 1, \dots, n$. Let $W = \sum_{i=1}^n I_i$ and $\lambda = \mathbb{E}W = \sum_{i=1}^n p_i$. For any choice of index $i \in \{1, \dots, n\}$, let $\Gamma_i^s \subseteq \{1, \dots, n\} \setminus \{i\}$ be the set of indices comprising all $j \neq i$ for which I_j is strongly dependent on I_i , and let Γ_i^w similarly be the set of indices j for which I_j is weakly dependent on I_i . Furthermore, let

$$W_i = \sum_{j=1, j \neq i}^n I_j, \quad Z_i = \sum_{j \in \Gamma_i^s} I_j, \quad \text{and } Y_i = \sum_{j \in \Gamma_i^w} I_j = W - I_i - Z_i = W_i - Z_i.$$

Then

$$d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda)) \leq \sum_{i=1}^n [(p_i^2 + p_i \mathbb{E} Z_i + \mathbb{E}(I_i Z_i))] \min\left(1, \frac{1}{\lambda}\right) + \sum_{i=1}^n \eta_i \min\left(1, \sqrt{\frac{2}{e\lambda}}\right),$$

where η_i is chosen such that

$$|\mathbb{E}[I_i g(Y_i + 1) - p_i \mathbb{E} g(Y_i + 1)]| \leq \eta_i \|g\|. \quad (2.20)$$

For instance, η_i may be chosen as follows:

$$\eta_i = \mathbb{E} |\mathbb{E}\{I_i | (I_j, j \in \Gamma_i^w)\} - p_i|.$$

Remark 2.4.7. In the case that Y_i is precisely independent of I_i the second error term $\sum_{i=1}^n \eta_i \min\{1, (2/e\lambda)^{1/2}\}$ disappears.

Proof. From (2.14), we have that

$$d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda)) = \sup_{A \subseteq \mathbb{Z}_+} |\mathbb{E}[\lambda g(W + 1) - W g(W)]|,$$

where $g = g_{\lambda, A}$. We replace Equation (2.17) by the following:

$$\begin{aligned} \mathbb{E}[I_i g(W)] &= \mathbb{E}[I_i g(W_i + 1)] \\ &= \mathbb{E}[I_i g(Y_i + 1)] + \mathbb{E}[I_i (g(Z_i + Y_i + 1) - g(Y_i + 1))]. \end{aligned}$$

Then,

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}[p_i g(W + 1) - I_i g(W)] &= \sum_{i=1}^n \{\mathbb{E}[p_i g(I_i + Z_i + Y_i + 1)] - \mathbb{E}[I_i g(Y_i + 1)] \\ &\quad - \mathbb{E}[I_i (g(Z_i + Y_i + 1) - g(Y_i + 1))]\}, \end{aligned}$$

and therefore, by adding and subtracting $p_i \mathbb{E}[g(Y_i + 1)]$,

$$\begin{aligned} &|\mathbb{E}[\lambda g(W + 1) - W g(W)]| \\ &\leq \sum_{i=1}^n |p_i \mathbb{E}[g(I_i + Z_i + Y_i + 1) - g(Y_i + 1)]| + |p_i \mathbb{E}[g(Y_i + 1)] - \mathbb{E}[I_i g(Y_i + 1)]| \\ &\quad + |\mathbb{E}[I_i (g(Z_i + Y_i + 1) - g(Y_i + 1))]|. \end{aligned} \quad (2.21)$$

With Lemma 2.4.3 (ii), we have $|g(j+k) - g(j)| \leq k\Delta g \leq k \min(1, \lambda^{-1})$. We thus find for the first and third terms in (2.21),

$$\begin{aligned} |p_i \mathbb{E}[g(I_i + Z_i + Y_i + 1) - g(Y_i + 1)]| &\leq p_i (\mathbb{E}I_i + \mathbb{E}Z_i) \min(1, \lambda^{-1}), \\ |\mathbb{E}[I_i(g(Z_i + Y_i + 1) - g(Y_i + 1))]| &\leq \mathbb{E}[I_i Z_i] \min(1, \lambda^{-1}), \end{aligned}$$

respectively. For the second term in (2.21) we choose η_i such that (2.20) is satisfied and use Lemma 2.4.3 (i). Finally, we show that (2.20) holds for the choice $\eta_i = \mathbb{E} |\mathbb{E} \{I_i | (I_j, j \in \Gamma_i^w)\} - p_i|$:

$$\begin{aligned} |\mathbb{E}[g(Y_i + 1)(p_i - I_i)]| &\leq \|g\| \cdot |\mathbb{E}[p_i - I_i]| \\ &= \|g\| \cdot |\mathbb{E}[\mathbb{E}[(p_i - I_i) | (I_j, j \in \Gamma_i^w)]]| \\ &\leq \eta_i \|g\|. \end{aligned}$$

□

2.4.4 The antisymmetric function approach

Stein's method for normal and for Poisson approximation can be put into a broader framework. A general approach using an exchangeable pair of random variables and an antisymmetric function was first proposed by Stein (1986) and later also discussed by Chen (1998) and Erhardsson (2005). We give a brief summary that closely follows Erhardsson (2005). Let (S, \mathcal{S}, μ) be a probability space, denote by \mathcal{H} the set of measurable functions $h : S \rightarrow \mathbb{R}$ and by $\mathcal{H}_0 \subset \mathcal{H}$ a set of μ -integrable functions. The goal is to compute $\int_S h d\mu$ for all $h \in \mathcal{H}_0$, but the structure of μ might be too complicated to do this. It could, for instance, be the distribution of a sum of a large number of dependent random variables. An idea to circumvent the problem of evaluating $\int_S h d\mu$ precisely is to instead replace μ by a probability measure μ_0 that is close to μ , with the advantage of being better known and easier to handle, classical examples for μ_0 being the normal and the Poisson distributions. So the new probability measure μ_0 should be chosen on (S, \mathcal{S}) such that the μ -integrable functions $h \in \mathcal{H}_0$ are also μ_0 -integrable and $\int_S h d\mu_0$ is easily computed for any $h \in \mathcal{H}_0$. It then remains to estimate (preferably uniformly over all $h \in \mathcal{H}_0$) the error of the approximation of $\int_S h d\mu$ by $\int_S h d\mu_0$. To that end, we have to find a set of functions \mathcal{G}_0 and a mapping $T_0 : \mathcal{G}_0 \rightarrow \mathcal{H}$ such that for all $h \in \mathcal{H}_0$, the equation

$$T_0 g = h - \int_S h d\mu_0,$$

that we call the *Stein equation*, has a solution $g \in \mathcal{G}_0$. We call T_0 a *Stein operator* for the distribution μ_0 . If the above equation holds, then

$$\int_S (T_0 g) d\mu = \int_S h d\mu - \int_S h d\mu_0.$$

The hope is then that it is easier to estimate $|\int_S (T_0 g) d\mu|$ than the actual approximation error $|\int_S h d\mu - \int_S h d\mu_0|$. But how to find a suitable operator T_0 ? Note that by the above Stein equation a necessary property for the Stein operator is that

$$\int_S (T_0 g) d\mu_0 = \int_S h d\mu_0 - \int_S h d\mu_0 = 0.$$

Stein (1986) proposed the following way to construct T_0 :

- (a) Let (Ω, \mathcal{F}, P) be a probability space with associated expectation operator \mathbb{E} . Let (X, Y) be an exchangeable pair of mappings of (Ω, \mathcal{F}, P) into the probability space (S, \mathcal{S}, μ_0) in the sense that $P(X \in A) = P(Y \in A) = \mu_0(A)$ for all $A \in \mathcal{S}$ and $P(X \in A, Y \in A') = P(X \in A', Y \in A)$ for all $A, A' \in \mathcal{S}$.
- (b) Choose a mapping $\alpha : \mathcal{G}_0 \rightarrow \mathcal{G}$, where \mathcal{G} is the space of antisymmetric functions $G : S^2 \rightarrow \mathbb{R}$ (i.e. $G(s, s') = -G(s', s)$ for all $s, s' \in S$) such that $\mathbb{E}|G(X, Y)| < \infty$.
- (c) Take $T_0 = T \circ \alpha$, where the linear mapping $T : \mathcal{G} \rightarrow \mathcal{H}$ is defined by $TG := \mathbb{E}^X G(X, Y)$, with \mathbb{E}^X denoting conditional expectation given X .

Then

$$\int_S (T_0 g) d\mu_0 = \int_S (TG) d\mu_0 = \mathbb{E} \mathbb{E}^X G(X, Y) = \mathbb{E} G(X, Y),$$

for all $g \in \mathcal{G}_0$, where $G := \alpha g$. Moreover,

$$\mathbb{E} G(X, Y) = \mathbb{E} G(Y, X) = \mathbb{E} [-G(X, Y)] = -\mathbb{E} G(X, Y),$$

where we use exchangeability for the first equality and the antisymmetry of G for the second. It follows that

$$\mathbb{E} G(X, Y) = \int_S (T_0 g) d\mu_0 = 0, \quad \text{for all } g \in \mathcal{G}_0, \quad (2.22)$$

and thus the necessary property for the Stein operator, called *Stein identity* for the target distribution μ_0 , is satisfied by the above choice of T_0 . The following subsection will go into further detail on how to apply this procedure for the example of approximation by a Poisson distribution. In general, there is unfortunately no guarantee that the procedure will yield a Stein operator T_0 giving sharp estimates for the approximation error. Additional considerations have to be made for each case.

2.4.5 Immigration-death processes and the generator interpretation

Barbour (1988) discovered a way to relate the Stein equation by Chen (1975a) to the generator of a Markov process whose equilibrium distribution is $\mu_0 = \text{Poi}(\lambda)$. This section gives a brief outline, while details will be discussed in a more general setting in Section 2.5.4 below. The Markov process in question is a stationary immigration-death process $Z := \{Z_t, t \in \mathbb{R}_+\}$ on \mathbb{Z}_+ with constant immigration rate $\lambda > 0$ and unit per-capita death rate, where Z_t describes the number of particles in a population at time t . For this process, immigrations of particles into the population and deaths of particles already in the population occur independently of one another. Also, each of the particles in the population dies after an $\text{Exp}(1)$ lifetime, independently of the others. As illustrated in Figure 2.1, when the population has size $k \in \mathbb{Z}_+$, i.e. when the process Z is in state k , transitions can only be to one of the adjoining states $k + 1$ (immigration of one particle with constant rate λ) or $k - 1$ (death of one particle with rate k , the current population size). Now remember from (2.13) that the Stein equation by Chen (1975a) is given by

$$\lambda g(k+1) - kg(k) = I_{\{k \in A\}} - \text{Poi}(\lambda)\{A\} \quad (=: f_A(k) =: f(k)), \quad \text{for all } k \in \mathbb{Z}_+.$$

This is a first-order equation. Barbour (1988) noted that if the solution g is written as the first backward difference $\nabla \gamma$ of a function γ , i.e. if $g(k) = \gamma(k) - \gamma(k-1)$ for all $k \in \mathbb{Z}_+$ (where $\gamma(-1) := 0$), the left-hand side of the Stein equation may be written as

$$\lambda g(k+1) - kg(k) = \lambda \gamma(k+1) + k \gamma(k-1) - (\lambda + k) \gamma(k) =: (\mathcal{A}\gamma)(k), \quad \text{for all } k \in \mathbb{Z}_+.$$

Here, \mathcal{A} is the infinitesimal generator of the process Z , and the Stein equation may now be reformulated by way of the second-order equation

$$\mathcal{A}\gamma = f. \tag{2.23}$$

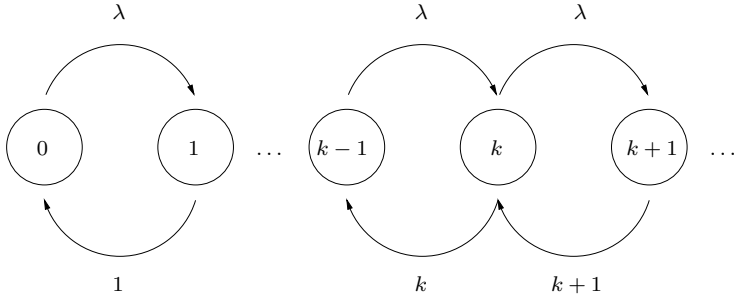


Figure 2.1: Transitions of the immigration-death process can only be to one of the neighbouring states. Immigration of one particle occurs with constant rate λ , whereas the death rate depends on the current population size as each particle in the population has unit per-capita death rate.

In order to determine γ , note that a solution x to an equation of the form $\mathcal{A}x = f$ is typically given by $x(z) = -\int_0^\infty \mathbb{E}^z f(Z_t) dt$, for all bounded f such that $\int_S f d\mu_0 = 0$, where Z is an immigration-death process on S with infinitesimal generator \mathcal{A} and equilibrium distribution μ_0 , and where \mathbb{E}^z denotes the distribution of the process conditioned on $Z_0 = z \in S$. In the case of the Stein equation (2.23), where the equilibrium distribution of the immigration-death process Z is $\mu_0 = \text{Poi}(\lambda)$ and the function $f(j) = I_{\{j \in A\}} - \text{Poi}(\lambda)\{A\}$ is obviously bounded for all $j \in S := \mathbb{Z}_+$, we indeed have

$$\int_S f d\mu_0 = \sum_{j \in \mathbb{Z}_+} [I_{\{j \in A\}} - \text{Poi}(\lambda)\{A\}] \text{Poi}(\lambda)\{j\} = \text{Poi}(\lambda)\{A\} - \text{Poi}(\lambda)\{A\} = 0,$$

and the solution γ of the Stein equation is thus given by

$$\begin{aligned} \gamma(k) &= -\int_0^\infty \mathbb{E}^k f(Z_t) dt \\ &= -\int_0^\infty \sum_{j \in \mathbb{Z}_+} [I_{\{j \in A\}} - \text{Poi}(\lambda)\{A\}] P(Z_t = j | Z_0 = k) dt \\ &= -\int_0^\infty [P(Z_t \in A | Z_0 = k) - \text{Poi}(\lambda)\{A\}] dt, \end{aligned}$$

for all $k \in \mathbb{Z}_+$ (see also, for instance, Theorem 2.4 in Erhardsson (2005)).

One of the advantages of the above approach is that it provides a probabilistic interpretation of the solution of the Stein equation, thus enabling the use of probabilistic arguments to determine smoothness estimates of the solution. Another advantage is that it is applicable to a wide range of approximation problems; most importantly for us, to the problem of approximating a point process by a Poisson process with the same mean measure. We refer to Section 2.5.4 for more details.

We now delineate the connection between the antisymmetric function approach from Section 2.4.4, applied to Poisson approximation, and the above generator interpretation. Let $(S, \mathcal{S}, \mu) = (\mathbb{Z}_+, \mathcal{P}(\mathbb{Z}_+), \mu)$, where $\mathcal{P}(\mathbb{Z}_+)$ is the power σ -algebra of \mathbb{Z}_+ . The aim is to approximate $\mu(A) = \int h d\mu$ by $\text{Poi}(\lambda)\{A\} = \int h d\mu_0$, where $h = I_A$, for any $A \in \mathcal{P}(\mathbb{Z}_+)$. Since the immigration-death process Z that we introduced above is reversible and has stationary distribution μ_0 , (Z_0, Z_t) is an exchangeable pair with marginal distribution μ_0 . Let $\mathcal{H} = \mathcal{H}$ be the set of real-valued functions on \mathbb{Z}_+ and let \mathcal{G} be the set of antisymmetric functions $\mathbb{Z}_+^2 \rightarrow \mathbb{R}$. Define $\alpha : \mathcal{H} \rightarrow \mathcal{G}$ by $(\alpha\gamma)(k, l) = \gamma(l) - \gamma(k)$ for functions γ that do not grow too fast and note that α is antisymmetric, since $\alpha\gamma(k, l) = -\alpha\gamma(l, k)$. For all $t \geq 0$, take $T_0^t = T_t \circ \alpha$, where we define $T_t : \mathcal{G} \rightarrow \mathcal{H}$ by $T_t G := \mathbb{E}^{Z_0} G(Z_0, Z_t)$. With $G := \alpha\gamma$, we then have

$$\mathbb{E}[T_0^t \gamma] = \mathbb{E}[T_t G] = \mathbb{E}G(Z_0, Z_t) = \mathbb{E}[\gamma(Z_t) - \gamma(Z_0)],$$

and, following the arguments from Section 2.4.4, we obtain $\mathbb{E}[T_0^t \gamma] = 0$ for all $t \geq 0$. In order to see the connection to the generator \mathcal{A} of the immigration-death process Z , note that for all $k \in \mathbb{Z}_+$,

$$\begin{aligned} (T_0^t \gamma)(k) &= \mathbb{E}[\alpha\gamma(Z_0, Z_t) | Z_0 = k] = \mathbb{E}[\gamma(Z_t) | Z_0 = k] - \gamma(k) \\ &= \sum_{l=k-1}^{k+1} \gamma(l) P(Z_t = l | Z_0 = k) - \gamma(k). \end{aligned}$$

As it can easily be shown that $P(Z_t = k+1 | Z_0 = k) = \lambda t + o(t)$, $P(Z_t = k-1 | Z_0 = k) = kt + o(t)$, and $P(Z_t = k | Z_0 = k) = 1 - (\lambda + k)t + o(t)$, we have that

$$\lim_{t \downarrow 0} \frac{1}{t} (T_0^t \gamma)(k) = \lambda\gamma(k+1) + k\gamma(k-1) - (\lambda + k)\gamma(k) = (\mathcal{A}\gamma)(k).$$

Informally (supposing that the limit and the expectation may be interchanged), we then observe the following connection between the Stein operators T_0^t and

the generator \mathcal{A} :

$$\lim_{t \downarrow 0} \frac{1}{t} \mathbb{E}[T_0^t \gamma] = 0 = \mathbb{E}[(\mathcal{A}\gamma)(Z_0)].$$

2.5 Poisson process approximation in the total variation distance

So far, we have been concerned with determining the error in total variation of the approximation of the law of a random variable $W = \sum_{i=1}^n I_i$ by a Poisson distribution with the same mean, the I_i 's being possibly non-identically distributed and/or dependent indicator variables. We may generalise W by instead considering point processes. Loosely speaking, these not only give a random number of “points” but also the random configuration of such points in space. As $\mathcal{L}(W)$ may be approximated by $\text{Poi}(\mathbb{E}W)$, we will show that the law $\mathcal{L}(\Xi)$ may similarly be approximated by the law of a Poisson point process with mean measure $\mathbb{E}\Xi$. Barbour and Brown (1992) and Barbour et al. (1992) extended the generator approach from the previous section to this problem. We formally define point processes in Section 2.5.1 and introduce the particular example of Poisson processes in Section 2.5.2. Section 2.5.3 gives short introductions to some tools from point process theory, namely Palm processes and Janossy densities, that we will need in Section 2.5.4 to study approximation errors in the total variation distance using the generator approach. Finally, Section 2.5.5 gives an easier way to determine approximation errors for the special case of marked point processes.

Throughout, let E be a locally compact separable metric space. In later applications, that is, starting from Chapter 3, we simply use $E \subseteq \mathbb{R}^d$, $d \geq 1$. Let E be equipped with its Borel σ -algebra $\mathcal{E} := \mathcal{B}(E)$, i.e. the σ -algebra generated by the open sets. The *Dirac measure* δ_z on \mathcal{E} for a point $z \in E$ is given by

$$\delta_z(B) = \begin{cases} 1 & \text{if } z \in B, \\ 0 & \text{if } z \notin B, \end{cases}$$

for any $B \in \mathcal{E}$. For a countable collection $\{z_i\}$, $i \geq 1$, of not necessarily distinct points in E , consider the counting measure $\xi := \sum_{i=1}^{\infty} \delta_{z_i}$ on \mathcal{E} , which assigns values in $\{0, 1, \dots\} \cup \{\infty\}$ to the sets that it measures. Suppose that ξ is a Radon measure, i.e. suppose that $\xi(K) < \infty$ for compact sets $K \in \mathcal{E}$. Integer-valued Radon measures such as ξ are called *point measures*, and sometimes also *point configurations* on E . Denote by $\overline{M}_p(E)$ the space of all point measures on

E and equip $\overline{M}_p(E)$ with the σ -algebra $\overline{\mathcal{M}}_p(E)$ that is the smallest σ -algebra containing all sets of the form $\{\xi \in \overline{M}_p(E) : \xi(B) \in M\}$ for any $B \in \mathcal{E}$ and for any Borel set $M \subset [0, \infty]$. In other words, $\overline{\mathcal{M}}_p(E)$ is the smallest σ -algebra making the evaluation maps $\xi \rightarrow \xi(B)$ from $\overline{M}_p(E)$ to $[0, \infty]$ measurable for any set $B \in \mathcal{E}$. Furthermore, denote by $M_p(E) \subset \overline{M}_p(E)$ the space of all *finite* point measures, i.e. of all point measures that assign values in $\{0, 1, \dots\}$ to the sets that they measure. Equip $M_p(E)$ with the σ -algebra $\mathcal{M}_p(E)$ that is the smallest σ -algebra making the evaluation maps $\xi \rightarrow \xi(B)$ from $M_p(E)$ to $[0, \infty)$ measurable for any set $B \in \mathcal{E}$.

2.5.1 Point process

Let (Ω, \mathcal{F}, P) be a probability space. A *point process* Ξ on E is a measurable map from a probability space to the space of point measures,

$$\begin{aligned} \Xi : (\Omega, \mathcal{F}, P) &\rightarrow (\overline{M}_p(E), \overline{\mathcal{M}}_p(E)) \\ \omega &\mapsto \Xi(\omega) = \xi. \end{aligned}$$

A point process is thus a random element of $\overline{M}_p(E)$, i.e., for fixed $\omega \in \Omega$, a realisation $\Xi(\omega) := \Xi(\omega, \cdot)$ is a point measure $\xi(\cdot) \in \overline{M}_p(E)$. For fixed $B \in \mathcal{E}$, $\Xi(B) := \Xi(\cdot, B)$ is a random variable taking values in $\{0, 1, \dots\} \cup \{\infty\}$. The space E that the point process lives on is called *state space*.

Example 2.5.1. There are numerous ways to represent a point process by way of the Dirac measure and random variables. Suppose, for instance, that Γ is a finite or countable index set, and that I_α , $\alpha \in \Gamma$, are i.i.d. Bernoulli random variables with probability of success $P(I_\alpha = 1) = p_\alpha \in (0, 1)$. Also, let X_α , $\alpha \in \Gamma$, be i.i.d. E -valued random variables, defined on the same probability space as the I_α 's, but independent of these. We give some examples of point processes:

- (a) $\sum_{\alpha \in \Gamma} I_\alpha \delta_\alpha$ is a point process with state space Γ .
- (b) For any integer $n \geq 1$ and for $\Gamma = \{1, \dots, n\}$, $\sum_{i=1}^n I_i \delta_{i/n}$ is a point process with state space $[0, 1]$.
- (c) $\sum_{\alpha \in \Gamma} I_\alpha \delta_{X_\alpha}$ is a point process with state space E . It is called *marked point process* and the X_α 's are called *marks*.

- (d) With $P(I_\alpha = 1) = 1$ in (c), for all $\alpha \in \Gamma$, the marked point process is $\Xi := \sum_{\alpha \in \Gamma} \delta_{X_\alpha}$. Here, $\Xi(\omega) = \sum_{\alpha \in \Gamma} \delta_{x_\alpha}$ gives the configuration of the points $x_\alpha = X_\alpha(\omega)$ in E , whereas $\Xi(B)$ gives the random number of points in the subset B of E . $\Xi(\omega, B)$ gives the number of points x_α lying in the set B ; see Figure 2.2 for an illustration.

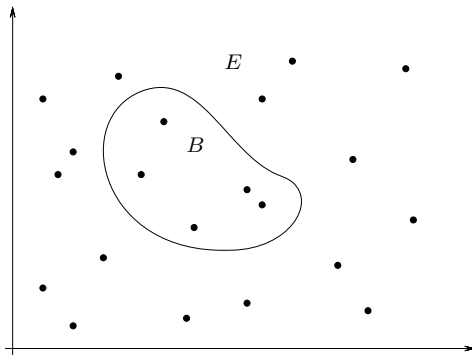


Figure 2.2: A configuration of points $x_\alpha = X_\alpha(\omega)$ in $E = \mathbb{R}_+^2$. The number of points falling into the set B is given by $\Xi(\omega, B)$.

The probability law P_Ξ of the point process Ξ is the measure $P \circ \Xi^{-1}(\cdot) = P(\Xi \in \cdot)$ on $\overline{\mathcal{M}}_p(E)$. It is uniquely determined by the set of *finite-dimensional distributions*, i.e. the distributions of random vectors $(\Xi(B_1), \dots, \Xi(B_m))$ for any choice of $m \geq 1$ and $B_1, \dots, B_m \in \mathcal{E}$. The *intensity measure* or *mean measure* of Ξ is the measure λ on \mathcal{E} defined, for any $B \in \mathcal{E}$, by

$$\lambda(B) = \mathbb{E}\Xi(B) = \int_{\Omega} \Xi(\omega, B) P(d\omega) = \int_{\overline{\mathcal{M}}_p(E)} \xi(B) P_\Xi(d\xi).$$

(Note that λ need not be Radon.) A point process Ξ is called *simple* if $\Xi(\omega, \{z\}) \leq 1$ for all $z \in E$ and for all $\omega \in \Omega$, i.e. if an element $z \in E$ can be hit by at most one point of the process. We call the point process Ξ *finite*, or more precisely, *almost surely finite*, if $P(|\Xi| < \infty) = 1$, where $|\Xi|$ denotes the random total number of points of Ξ .

2.5.2 Poisson process

Let λ be a σ -finite measure on (E, \mathcal{E}) . By this we mean that E may be written as a countable union of sets, each of which has a finite measure. It also implies that λ is locally finite, i.e. every point $z \in E$ has a neighbourhood with finite measure. An example of a σ -finite measure is the Lebesgue measure on Euclidean spaces.

A point process Ξ on E is called a *Poisson process* or *Poisson random measure* (PRM) with mean measure λ if Ξ satisfies:

- (a) For any $B \in \mathcal{E}$, we have that $\Xi(B) \sim \text{Poi}(\lambda(B))$, i.e.

$$P(\Xi(B) = k) = \begin{cases} \frac{(\lambda(B))^k}{k!} e^{-\lambda(B)}, & \lambda(B) < \infty, \\ 0, & \lambda(B) = \infty, \end{cases} \quad \text{for any } k \in \mathbb{Z}_+.$$

- (b) For any $m \geq 1$, if B_1, \dots, B_m are mutually disjoint sets in \mathcal{E} , then $\Xi(B_1), \dots, \Xi(B_m)$ are independent random variables.

It follows from (a) that $\lambda(B) = \infty$ implies $P(\Xi(B) = \infty) = 1$. By Proposition 3.6(i) in Resnick (1987) we know that, given a σ -finite measure λ , a Poisson process with mean measure λ exists and its law is uniquely determined by (a) and (b). We denote the law of a Poisson process Ξ with mean measure λ by $\text{PRM}(\lambda)$, i.e. $\Xi \sim \text{PRM}(\lambda)$.

Example 2.5.2. Suppose that $E \subseteq \mathbb{R}^d$, $d \geq 1$.

- (a) Let $\lambda = \lambda|\cdot|$, where $\lambda > 0$ and $|\cdot|$ denotes Lebesgue measure on E . Then $\Xi \sim \text{PRM}(\lambda|\cdot|)$ is called a *homogeneous Poisson process with intensity λ* .
- (b) Alternatively, suppose that the mean measure λ of a Poisson process Ξ is absolutely continuous with respect to Lebesgue measure, i.e. that there exists a non-negative function λ such that for any $B \in \mathcal{E}$,

$$\lambda(B) = \int_B \lambda(x) dx.$$

Then Ξ is called *inhomogeneous Poisson process* with *rate* or *intensity function* $\lambda(\cdot)$.

2.5.3 Palm processes and Janossy densities

Suppose Ξ is a point process on E with σ -finite mean measure λ . For any $z \in E$, a point process Ξ_z is called *Palm process associated with Ξ at z* if, for any measurable function $f : E \times \overline{M}_p(E) \rightarrow \mathbb{R}_+$,

$$\mathbb{E} \left[\int_E f(z, \Xi) \Xi(dz) \right] = \mathbb{E} \left[\int_E f(z, \Xi_z) \lambda(dz) \right]. \quad (2.24)$$

We may define probability measures $\{P_z, z \in E\}$, called *Palm distributions*, by setting

$$P_z(R) := P(\Xi_z \in R) := \frac{\mathbb{E} [I_{\{\Xi \in R\}} \Xi(dz)]}{\lambda(dz)},$$

for all $R \in \overline{M}_p(E)$. A point process Ξ_z on E is then called a *Palm process associated with Ξ at z* if it has the Palm distribution P_z of Ξ at z . Palm processes can be used to give a characterisation of Poisson processes:

A process Ξ is a Poisson process if and only if $\mathcal{L}(\Xi_z) = \mathcal{L}(\Xi + \delta_z)$ λ -a.s. (2.25)

For more details on Palm theory, see Chapter 10 in Kallenberg (1986) or Chapter 13 in Daley and Vere-Jones (2008).

Another important tool in point process theory is given by the so-called Janossy measures. These are used to express the probability of a point process having a certain number of points and these points being located in a certain region. Suppose that Ξ is a finite point process on E . Then there exist measures $\{J_m\}$, $m \geq 0$, called *Janossy measures*, such that for measurable functions $f : M_p(E) \rightarrow \mathbb{R}_+$,

$$\mathbb{E} f(\Xi) = \sum_{m=0}^{\infty} \int_{E^m} f \left(\sum_{i=1}^m \delta_{z_i} \right) \frac{1}{m!} J_m(dz_1, \dots, dz_m),$$

where $J_m(dz_1, \dots, dz_m)/m!$ describes the probability of the process having m points lying close to z_1, \dots, z_m . Suppose there exists a fixed σ -finite measure ν on E . For instance, let ν be the counting measure in case E is a finite set, or let it be Lebesgue measure for E a compact subset of Euclidean space. Suppose furthermore that for each $m \geq 0$, J_m is absolutely continuous with respect to ν . The Radon-Nikodym theorem then ensures the existence of derivatives

$j_m : E^m \rightarrow [0, \infty)$ of J_m with respect to $\boldsymbol{\nu}^m$, so that

$$\mathbb{E}f(\Xi) = \sum_{m \geq 0} \int_{E^m} f \left(\sum_{i=1}^m \delta_{z_i} \right) \frac{1}{m!} j_m(z_1, \dots, z_m) \boldsymbol{\nu}^m(dz_1, \dots, dz_m).$$

The derivatives $\{j_m\}$, $m \geq 0$, are called *Janossy densities*. In the above expression for $\mathbb{E}f(\Xi)$, the term with $m = 0$ is interpreted as $j_0 f(\emptyset)$. By Lemma 5.4.III in Daley and Vere-Jones (2003), the density μ of the first moment measure of Ξ , i.e. of the intensity measure $\boldsymbol{\lambda}$ of Ξ , may then be expressed in terms of the Janossy densities:

$$\mu(z) = \sum_{m \geq 0} \int_{E^m} \frac{1}{m!} j_{m+1}(z, z_1, \dots, z_m) \boldsymbol{\nu}^m(dz_1, \dots, dz_m), \quad (2.26)$$

where the term with $m = 0$ is interpreted as $j_1(z)$. We then have $\boldsymbol{\lambda}(dz) = \mu(z) \boldsymbol{\nu}(dz)$. Janossy densities may furthermore be used to express the conditional probability density of a point being located at z given the configuration Ξ^z of Ξ outside a neighbourhood N_z of z . More precisely, suppose that the point process Ξ is simple and that for each $z \in E$, $N_z \in \mathcal{E}$ is a neighbourhood of z , with $z \in N_z$, such that the following two mappings are product measurable:

$$\begin{aligned} \mathcal{M}_p(E) \times E &\rightarrow [0, \infty) : (\xi, z) \mapsto \xi(N_z), \\ \mathcal{M}_p(E) \times E &\rightarrow \mathcal{M}_p(E) : (\xi, z) \mapsto \xi \text{ restricted to } N_z^c. \end{aligned} \quad (2.27)$$

For any $z \in E$ and for some fixed integer m , fix m points $x_1, \dots, x_m \in N_z^c$ and let $\mathbf{x} = (x_1, \dots, x_m)$. Define

$$g(z, \mathbf{x}) = \frac{\sum_{k \geq 0} \int_{N_z^k} j_{m+k+1}(z, \mathbf{x}, \mathbf{y}) (k!)^{-1} \boldsymbol{\nu}^k(d\mathbf{y})}{\sum_{l \geq 0} \int_{N_z^l} j_{m+l}(\mathbf{x}, \mathbf{w}) (l!)^{-1} \boldsymbol{\nu}^l(d\mathbf{w})}. \quad (2.28)$$

$g(z, \mathbf{x})$ is the conditional density of a point at z given that Ξ^z is $\sum_{i=1}^m \delta_{x_i}$ (the term with $k = 0$ is interpreted as $j_{m+1}(z, \mathbf{x})$ and the term with $l = 0$ similarly; moreover, if the denominator is zero, we interpret $g(z, \mathbf{x})$ as zero.) Barbour and Brown (1992) (see (2.7) on p. 16) show that, for a bounded measurable function $f : \mathcal{M}_p(E) \rightarrow \mathbb{R}_+$, we then have

$$\mathbb{E} \left[\int_E f(\Xi^z) \Xi(dz) \right] = \mathbb{E} \left[\int_E f(\Xi^z) g(z, \Xi^z) \boldsymbol{\nu}(dz) \right]. \quad (2.29)$$

Finally, like Palm processes, Janossy densities may also be used to give a characterisation of Poisson processes (see Theorem 2.11 in Xia (2005)):

Theorem 2.5.3. *A point process Ξ on E with mean measure λ and $\lambda = \lambda(E) < \infty$ is a Poisson process if and only if, with respect to λ , its Janossy densities j_m are constant and equal to $e^{-\lambda}$, for all $m \in \mathbb{Z}_+$.*

For more details on Janossy measures and densities, consult Sections 5.3 and 5.4 in Daley and Vere-Jones (2003).

2.5.4 Approximation of point processes – the generator interpretation

Let Ξ be a finite point process on E with finite intensity measure λ , where $\lambda := \lambda(E) < \infty$. Let $Z := \{Z_t, t \in \mathbb{R}_+\}$ be an immigration-death process on E with immigration intensity λ and unit per-capita death rate. This process is called *spatial immigration-death process* by Preston (1975) and Xia (2005). Z_t takes values in $M_p(E)$ and describes the point configuration of particles of a population on E at time t . Given that the process takes a configuration $\xi \in M_p(E)$, the process stays in state ξ for an $\text{Exp}(1/(|\xi| + \lambda))$ -distributed period of time. Then, with probability $\lambda/(|\xi| + \lambda)$, a new particle immigrates to the population and puts itself on $z \in E$, which is chosen from the distribution λ/λ , independently of the existing configuration. The new configuration is then $\xi + \delta_z$. Or, with probability $|\xi|/(|\xi| + \lambda)$, one particle from the population dies, that is, a point, say, δ_w , is chosen uniformly at random from the existing configuration and is erased. The new configuration is then $\xi - \delta_w$. See Figure 2.3 for an illustration.

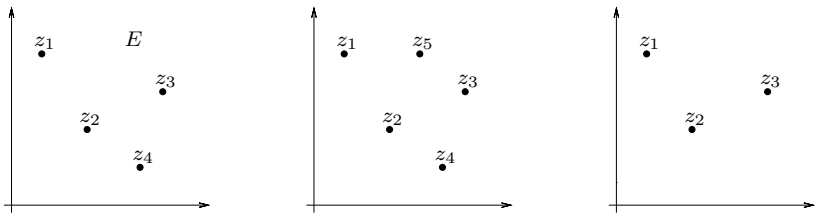


Figure 2.3: (Left) A configuration ξ of four points z_1, z_2, z_3, z_4 in $E = \mathbb{R}_+^2$. (Middle) An additional point z_5 is added to ξ so that the new configuration is $\xi + \delta_{z_5}$. (Right) Point z_4 disappears from ξ . The new configuration is $\xi - \delta_{z_4}$.

The generator of Z is given by

$$\begin{aligned}
 & (\mathcal{A}\gamma)(\xi) \\
 &= \lim_{t \downarrow 0} \frac{1}{t} \{ \mathbb{E} [\gamma(Z_t) | Z_0 = \xi] - \gamma(\xi) \} \\
 &= \lim_{t \downarrow 0} \frac{1}{t} \left\{ \int_E \sum_{\zeta \in \{\xi + \delta_z, \xi - \delta_z, \xi\}} \gamma(\zeta) P(Z_t = \zeta | Z_0 = \xi) dz - \gamma(\xi) \right\} \\
 &= \lim_{t \downarrow 0} \frac{1}{t} \left\{ \int_E \gamma(\xi + \delta_z) t \boldsymbol{\lambda}(dz) + \int_E \gamma(\xi - \delta_z) t \xi(dz) \right. \\
 &\quad \left. + \gamma(\xi) \left[1 - \int_E t(\boldsymbol{\lambda} + \xi)(dz) \right] - \gamma(\xi) \right\} \\
 &= \int_E [\gamma(\xi + \delta_z) - \gamma(\xi)] \boldsymbol{\lambda}(dz) + \int_E [\gamma(\xi - \delta_z) - \gamma(\xi)] \xi(dz), \tag{2.30}
 \end{aligned}$$

for all $\xi \in M_p(E)$ and for γ a suitable function $M_p(E) \rightarrow \mathbb{R}$; see also (3.6) in Xia (2005). Let $\Xi_{\boldsymbol{\lambda}}$ be a Poisson process with intensity measure $\boldsymbol{\lambda}$, i.e. $\Xi_{\boldsymbol{\lambda}} \sim \text{PRM}(\boldsymbol{\lambda})$. By (2.30),

$$\begin{aligned}
 & \mathbb{E} [(\mathcal{A}\gamma)(\Xi_{\boldsymbol{\lambda}})] \\
 &= \mathbb{E} \left\{ \int_E [\gamma(\Xi_{\boldsymbol{\lambda}} + \delta_z) - \gamma(\Xi_{\boldsymbol{\lambda}})] \boldsymbol{\lambda}(dz) + \int_E [\gamma(\Xi_{\boldsymbol{\lambda}} - \delta_z) - \gamma(\Xi_{\boldsymbol{\lambda}})] \Xi_{\boldsymbol{\lambda}}(dz) \right\}.
 \end{aligned}$$

For any $z \in E$, let $\Xi_{\boldsymbol{\lambda},z}$ be the Palm process associated with $\Xi_{\boldsymbol{\lambda}}$ at z . Then, by (2.24) and (2.25),

$$\begin{aligned}
 \mathbb{E} \int_E [\gamma(\Xi_{\boldsymbol{\lambda}} - \delta_z) - \gamma(\Xi_{\boldsymbol{\lambda}})] \Xi_{\boldsymbol{\lambda}}(dz) &= \mathbb{E} \int_E [\gamma(\Xi_{\boldsymbol{\lambda},z} - \delta_z) - \gamma(\Xi_{\boldsymbol{\lambda},z})] \boldsymbol{\lambda}(dz) \\
 &= \mathbb{E} \int_E [\gamma(\Xi_{\boldsymbol{\lambda}}) - \gamma(\Xi_{\boldsymbol{\lambda}} + \delta_z)] \boldsymbol{\lambda}(dz),
 \end{aligned}$$

and it follows that

$$\begin{aligned}
 \mathbb{E} [(\mathcal{A}\gamma)(\Xi_{\boldsymbol{\lambda}})] &= \mathbb{E} \int_E [\gamma(\Xi_{\boldsymbol{\lambda}} + \delta_z) - \gamma(\Xi_{\boldsymbol{\lambda}})] \boldsymbol{\lambda}(dz) \\
 &\quad + \mathbb{E} \int_E [\gamma(\Xi_{\boldsymbol{\lambda}}) - \gamma(\Xi_{\boldsymbol{\lambda}} + \delta_z)] \boldsymbol{\lambda}(dz) = 0.
 \end{aligned} \tag{2.31}$$

Together with Theorem 7.1 in Preston (1975), this implies that $\text{PRM}(\boldsymbol{\lambda})$ is the unique equilibrium distribution of Z (see also Proposition 3.4 in Xia (2005)). We next fix some notation:

- $\text{PRM}(\boldsymbol{\lambda})\{R\} := P_{\Xi_{\boldsymbol{\lambda}}}(R) = P(\Xi_{\boldsymbol{\lambda}} \in R)$, for any $R \in \mathcal{M}_p(E)$.
- $\text{PRM}(\boldsymbol{\lambda})(h) := \int_{M_p(E)} h(\xi) \text{PRM}(\boldsymbol{\lambda})\{d\xi\}$, for a function $h : M_p(E) \rightarrow \mathbb{R}$.
- For a fixed set $R \in \mathcal{M}_p(E)$ and the particular choice $h(\xi) := I_{\{\xi \in R\}}$, we then have

$$\text{PRM}(\boldsymbol{\lambda})(h) = \text{PRM}(\boldsymbol{\lambda})\{R\}.$$

- Let \mathbb{P}^ξ denote the distribution of the immigration-death process Z given that it has the initial configuration $\xi \in M_p(E)$, i.e.

$$\mathbb{P}^\xi(Z_t = \zeta) = P(Z_t = \zeta | Z_0 = \xi),$$

for any $\zeta \in M_p(E)$ and $t \geq 0$. Likewise, $\mathbb{E}^\xi f(Z_t) = \int_{M_p(E)} f(\zeta) P(Z_t = \zeta | Z_0 = \xi) d\zeta$.

The process analogue of the Stein equation is given by

$$(\mathcal{A}\gamma)(\xi) = h(\xi) - \text{PRM}(\boldsymbol{\lambda})(h), \quad (2.32)$$

for any $\xi \in M_p(E)$. We construct a solution γ to this equation: by Proposition 2.5.5 below, the function γ , given in Proposition 2.5.4, solves the Stein equation.

Proposition 2.5.4. *Let $Z = \{Z_t, t \in \mathbb{R}_+\}$ be an immigration-death process on E with immigration intensity $\boldsymbol{\lambda}$ and unit per-capita death rate. For any bounded $h : M_p(E) \rightarrow \mathbb{R}$, the function $\gamma : M_p(E) \rightarrow \mathbb{R}$ given by*

$$\gamma(\xi) = - \int_0^\infty \{ \mathbb{E}^\xi h(Z_t) - \text{PRM}(\boldsymbol{\lambda})(h) \} dt$$

is well defined, and $\sup_{\xi: \xi(E)=k} |\gamma(\xi)| < \infty$ for each $k \in \mathbb{Z}_+$.

Proof. We consider a coupling of the immigration-death process Z under \mathbb{P}^ξ with another immigration-death process \tilde{Z} under $\mathbb{P}^{\text{PRM}(\boldsymbol{\lambda})}$, setting $Z = Z^0 + D$, $\tilde{Z} = Z^0 + \tilde{D}$, where Z^0 , D and \tilde{D} are independent, Z^0 is an immigration-death process under \mathbb{P}^0 , i.e. having no initial particles, and D and \tilde{D} are

both pure death processes with unit per-capita death rate, such that $D_0 = \xi$, and $\tilde{D}_0 \sim \text{PRM}(\lambda)$. Let τ denote the earliest time at which both pure death processes have lost all of their particles:

$$\tau = \inf\{u \geq 0 : D_u = \tilde{D}_u = 0\}.$$

After time τ , the two processes Z and \tilde{Z} behave identically, i.e. $Z_t = \tilde{Z}_t$ for all $t \geq \tau$. Then,

$$\begin{aligned} |\gamma(\xi)| &\leq \int_0^\infty |\mathbb{E}^\xi h(Z_t) - \text{PRM}(\lambda)(h)| dt = \int_0^\infty |\mathbb{E}^\xi h(Z_t) - \mathbb{E}h(\tilde{Z}_t)| dt \\ &\leq \int_0^\infty \mathbb{E} \left(|h(Z_t) - h(\tilde{Z}_t)| \cdot I_{\{\tau > t\}} \mid D_0 = \xi \right) dt \\ &\leq 2\|h\| \int_0^\infty P(\tau > t \mid D_0 = \xi) dt = 2\|h\| \mathbb{E}(\tau \mid D_0 = \xi). \end{aligned} \quad (2.33)$$

To determine $\mathbb{E}(\tau \mid D_0 = \xi)$, note that the total number of points that the two processes D and \tilde{D} have to lose until time τ is given by the random integer $|D_0| + |\tilde{D}_0|$, and that

$$\tau = \sum_{i=1}^{|D_0|+|\tilde{D}_0|} \tau_i,$$

where τ_i denotes the time between the $(i-1)$ th and i th death. Since the two pure death processes have unit per-capita death rates, the time τ_i , for a fixed realisation $k = |\xi| + |\tilde{D}_0|(\omega)$, is exponentially distributed with rate $k - i + 1$, for each $i = 1, \dots, k$. Therefore,

$$|\gamma(\xi)| \leq 2\|h\| \mathbb{E}\psi(|\xi| + |\tilde{D}_0|) < \infty,$$

where $\psi(k) = \sum_{i=1}^k 1/(k - i + 1) = \sum_{i=1}^k 1/i$, and the last inequality follows because $|\tilde{D}_0| \sim \text{Poi}(\lambda)$. \square

Proposition 2.5.5. *The function γ defined in Proposition 2.5.4 satisfies the Stein equation*

$$(\mathcal{A}\gamma)(\xi) = h(\xi) - \text{PRM}(\lambda)(h),$$

for all $\xi \in M_p(E)$.

Proof. Let $\gamma_t(\xi) = -\int_0^t \{\mathbb{E}^\xi h(Z_u) - \text{PRM}(\boldsymbol{\lambda})(h)\} du$, and let T denote an exponential random variable modelling the first time that a particle is born or dies, with rate $q_\xi = \lambda + |\xi|$. Then, we may rewrite $\gamma_t(\xi)$ as

$$\begin{aligned}
 & - \int_0^t \left\{ \mathbb{E}^\xi [h(Z_u) - \text{PRM}(\boldsymbol{\lambda})(h)] \cdot I_{\{T > t\}} \right. \\
 & \quad \left. + \mathbb{E}^\xi [h(Z_u) - \text{PRM}(\boldsymbol{\lambda})(h)] \cdot I_{\{T \leq t\}} \right\} du \\
 &= -[h(\xi) - \text{PRM}(\boldsymbol{\lambda})(h)]e^{-q_\xi t} + \int_0^t e^{-q_\xi u} \left\{ -q_\xi u [h(\xi) - \text{PRM}(\boldsymbol{\lambda})(h)] \right. \\
 & \quad \left. + \int_E \gamma_{t-u}(\xi + \delta_z) \boldsymbol{\lambda}(dz) + \int_E \gamma_{t-u}(\xi - \delta_z) \xi(dz) \right\} du \\
 &= -\frac{1}{q_\xi} [h(\xi) - \text{PRM}(\boldsymbol{\lambda})(h)] (1 - e^{-q_\xi t}) \\
 & \quad + \int_0^t e^{-q_\xi u} \left\{ \int_E \gamma_{t-u}(\xi + \delta_z) \boldsymbol{\lambda}(dz) + \int_E \gamma_{t-u}(\xi - \delta_z) \xi(dz) \right\} du,
 \end{aligned}$$

where we obtained the last equality by evaluating the integral $-\int_0^t e^{-q_\xi u} q_\xi u [h(\xi) - \text{PRM}(\boldsymbol{\lambda})(h)] du$. From the proof of Proposition 2.5.4, the functions $\gamma_t(\xi)$ are uniformly bounded in t for each ξ . Letting $t \rightarrow \infty$ and using dominated convergence, it follows that

$$\begin{aligned}
 \gamma(\xi) &= -\frac{1}{q_\xi} [h(\xi) - \text{PRM}(\boldsymbol{\lambda})(h)] \\
 & \quad + \int_0^\infty e^{-q_\xi u} \left\{ \int_E \gamma(\xi + \delta_z) \boldsymbol{\lambda}(dz) + \int_E \gamma(\xi - \delta_z) \xi(dz) \right\} du \\
 &= \frac{1}{q_\xi} \left\{ -[h(\xi) - \text{PRM}(\boldsymbol{\lambda})(h)] + \int_E \gamma(\xi + \delta_z) \boldsymbol{\lambda}(dz) + \int_E \gamma(\xi - \delta_z) \xi(dz) \right\}
 \end{aligned} \tag{2.34}$$

By rearranging (2.34) and noting that $-q_\xi \gamma(\xi) = -\int_E \gamma(\xi)(\boldsymbol{\lambda}(dz) + \xi(dz))$, we find

$$\begin{aligned}
 & h(\xi) - \text{PRM}(\boldsymbol{\lambda})(h) \\
 &= \int_E [\gamma(\xi + \delta_z) - \gamma(\xi)] \boldsymbol{\lambda}(dz) + \int_E [\gamma(\xi - \delta_z) - \gamma(\xi)] \xi(dz) \\
 &= (\mathcal{A}\gamma)(\xi).
 \end{aligned}$$

□

The solution γ from Proposition 2.5.4 solves the Stein equation for any bounded function $h : M_p(E) \rightarrow \mathbb{R}$, and by (2.31),

$$\mathbb{E}[(\mathcal{A}\gamma)(\Xi_\lambda)] = 0,$$

as there is no expected variation of the immigration-death process Z when it is in equilibrium. In order to determine approximation errors in the total variation distance we now choose $h(\xi) = I_{\{\xi \in R\}}$ for some set $R \in \mathcal{M}_p(E)$. Then, taking expectations on both sides of the Stein equation (2.32) and replacing ξ by the process Ξ , we obtain

$$|\mathbb{E}(\mathcal{A}\gamma)(\Xi)| = |\mathbb{E}I_{\{\Xi \in R\}} - \text{PRM}(\lambda)\{R\}| = |P(\Xi \in R) - P(\Xi_\lambda \in R)|, \quad (2.35)$$

which is obviously 0 for $\Xi = \Xi_\lambda$, showing that the generator \mathcal{A} satisfies the required Stein identity (mentioned previously in Sections 2.4.4 and 2.4.5). Then the problem of determining an upper bound on the total variation distance between the laws of the two processes Ξ and $\Xi_\lambda \sim \text{PRM}(\lambda)$, that is,

$$d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\lambda)) = \sup_{R \in \mathcal{M}_p(E)} |P(\Xi \in R) - P(\Xi_\lambda \in R)|,$$

is equivalent to determining a uniform bound on $|\mathbb{E}(\mathcal{A}\gamma)(\Xi)|$. To achieve the latter, we require smoothness estimates of the function γ :

Lemma 2.5.6. *If γ is defined as in Proposition 2.5.4 and if $h(\xi) = I_{\{\xi \in R\}}$ for some set $R \in \mathcal{M}_p(E)$, then*

$$\begin{aligned} (i) \quad \Delta_1 \gamma &= \sup_{\xi \in M_p(E), z \in E} |\gamma(\xi + \delta_z) - \gamma(\xi)| \leq 1, \\ (ii) \quad \Delta_2 \gamma &= \sup_{\xi \in M_p(E); z, w \in E} |\gamma(\xi + \delta_z + \delta_w) - \gamma(\xi + \delta_z) - \gamma(\xi + \delta_w) + \gamma(\xi)| \\ &\leq 1. \end{aligned}$$

Proof. (i) From the definition of γ in Proposition 2.5.4 we have that for any $\xi \in M_p(E)$ and for any $z \in E$,

$$\gamma(\xi + \delta_z) - \gamma(\xi) = \int_0^\infty \{\mathbb{E}^\xi h(Z_t) - \mathbb{E}^{\xi + \delta_z} h(Z_t)\} dt,$$

where the immigration-death process $Z = \{Z_t, t \in \mathbb{R}_+\}$ on E with immigration intensity λ and unit per-capita death rate is realised under \mathbb{P}^ξ . Let T be an exponential random variable with rate 1 that is independent of Z . It follows that the process $Z'_t = Z_t + \delta_z I_{\{T > t\}}$ has distribution $\mathbb{P}^{\xi + \delta_z}$. Then

$$\begin{aligned} \gamma(\xi + \delta_z) - \gamma(\xi) &= \int_0^\infty \mathbb{E}^\xi [\{h(Z_t) - h(Z'_t)\} I_{\{T > t\}}] dt \\ &= \int_0^\infty \mathbb{E}^\xi [h(Z_t) - h(Z_t + \delta_z)] e^{-t} dt \end{aligned}$$

Since $\int_0^\infty e^{-t} dt = 1$ and

$$|h(\xi) - h(\xi + \delta_z)| = |I_{\{\xi \in R\}} - I_{\{\xi + \delta_z \in R\}}| \leq 1,$$

we have $\Delta_1 \gamma \leq 1$. (ii) By the definition of γ , we have

$$\begin{aligned} &\gamma(\xi + \delta_z + \delta_w) - \gamma(\xi + \delta_z) - \gamma(\xi + \delta_w) + \gamma(\xi) \\ &= - \int_0^\infty \{\mathbb{E}^{\xi + \delta_z + \delta_w} h(Z_t) - \mathbb{E}^{\xi + \delta_z} h(Z_t) - \mathbb{E}^{\xi + \delta_w} h(Z_t) + \mathbb{E}^\xi h(Z_t)\} dt. \end{aligned}$$

Let Z be realised under \mathbb{P}^ξ as in (i) and let T^z and T^w be two independent exponential random variables with rate 1. The processes

$$\begin{aligned} Z_t^z &= Z_t + \delta_z I_{\{T^z > t\}}, \\ Z_t^w &= Z_t + \delta_w I_{\{T^w > t\}}, \\ Z_t^{zw} &= Z_t + \delta_z I_{\{T^z > t\}} + \delta_w I_{\{T^w > t\}} \end{aligned}$$

then have distributions $\mathbb{P}^{\xi + \delta_z}$, $\mathbb{P}^{\xi + \delta_w}$ and $\mathbb{P}^{\xi + \delta_z + \delta_w}$, respectively. Therefore,

$$\begin{aligned} &\gamma(\xi + \delta_z + \delta_w) - \gamma(\xi + \delta_z) - \gamma(\xi + \delta_w) + \gamma(\xi) \\ &= - \int_0^\infty \mathbb{E}^\xi [\{h(Z_t^{zw}) - h(Z_t^z) - h(Z_t^w) + h(Z_t)\} I_{\{T^z > t\}} I_{\{T^w > t\}}] dt \\ &= - \int_0^\infty \mathbb{E}^\xi [h(Z_t + \delta_z + \delta_w) - h(Z_t + \delta_z) - h(Z_t + \delta_w) + h(Z_t)] e^{-2t} dt, \end{aligned}$$

which gives $\Delta_2 \gamma \leq 1$. □

We are now in shape to prove a process analogue of Theorem 2.4.6:

Theorem 2.5.7. *Suppose there exists a fixed measure ν on E and suppose that Ξ is a finite simple point process on E with finite mean measure λ and Janossy densities $\{j_m\}_{m \geq 0}$. Let the density μ of λ with respect to ν be given by (2.26) and let $\{N_z\}_{z \in E}$ be a neighbourhood structure satisfying (2.27). Then,*

$$\begin{aligned} d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\lambda)) &\leq \int_E \mathbb{E} \Xi(N_z) \mu(z) \nu(dz) + \mathbb{E} \left[\int_E \Xi(N_z \setminus \{z\}) \Xi(dz) \right] \\ &\quad + \int_E \mathbb{E} |g(z, \Xi^z) - \mu(z)| \nu(dz), \end{aligned}$$

where the conditional density $g(z, \Xi^z)$ at z given the configuration Ξ^z of Ξ outside N_z is defined in (2.28).

Remark 2.5.8. Since E is a metric space, an example for a neighbourhood N_z would be a closed ball with a certain radius centred at z . Barbour and Brown (1992) show that this choice satisfies (2.27).

Remark 2.5.9. Intuitively, the first error term in Theorem 2.5.7 measures the size of the neighbourhoods, the second measures the extent of local dependence, i.e. inside a neighbourhood, whereas the third term measures the size of the difference between what happens at z and what happens outside its neighbourhood N_z . There is clearly a trade-off between the sizes of the first and third error terms – the smaller N_z , the bigger the dependence between z and N_z^c , and vice versa.

Proof. From (2.35) we know that it is sufficient to find a uniform bound for the modulus of $\mathbb{E}(\mathcal{A}\gamma)(\Xi)$, where \mathcal{A} is defined as in (2.30), and γ is the solution of the Stein equation from Proposition 2.5.4 with $h(\xi) = I_{\{\xi \in R\}}$, for any $R \in \mathcal{M}_p(E)$. We have

$$\begin{aligned} &|\mathbb{E}(\mathcal{A}\gamma)(\Xi)| \\ &= \left| \mathbb{E} \left\{ \int_E [\gamma(\Xi + \delta_z) - \gamma(\Xi)] \lambda(dz) + \int_E [\gamma(\Xi - \delta_z) - \gamma(\Xi)] \Xi(dz) \right\} \right|. \end{aligned}$$

Let Ξ^z denote the configuration of Ξ outside N_z . We add and subtract $\gamma(\Xi^z +$

$\delta_z) - \gamma(\Xi^z)$ to both of the integrands above. Then,

$$\begin{aligned} |\mathbb{E}(\mathcal{A}\gamma)(\Xi)| &\leq \left| \mathbb{E} \int_E [\gamma(\Xi + \delta_z) - \gamma(\Xi) - \gamma(\Xi^z + \delta_z) + \gamma(\Xi^z)] \lambda(dz) \right| \\ &\quad + \left| \mathbb{E} \int_E [\gamma(\Xi) - \gamma(\Xi^z + \delta_z) - \gamma(\Xi - \delta_z) + \gamma(\Xi^z)] \Xi(dz) \right| \quad (2.36) \\ &\quad + \left| \mathbb{E} \int_E [\gamma(\Xi^z + \delta_z) - \gamma(\Xi^z)] \{\Xi(dz) - \lambda(dz)\} \right|. \end{aligned}$$

Denote realisations of Ξ and Ξ^z by ξ and ξ^z , respectively, and note that

$$\xi = \xi^z + \sum_{w \in N_z} \delta_w = \xi^z + \sum_{w \in N_z \setminus \{z\}} \delta_w + \delta_z,$$

where the last equality holds only for $z \in E$ such that $\Xi(\{z\}) = 1$. Then the modulus of the integrand of the first summand in (2.36) corresponds to

$$\begin{aligned} \left| \gamma \left(\xi^z + \sum_{w \in N_z} \delta_w + \delta_z \right) - \gamma \left(\xi^z + \sum_{w \in N_z} \delta_w \right) - \gamma(\xi^z + \delta_z) + \gamma(\xi^z) \right| \\ \leq \xi(N_z) \Delta_2 \gamma, \end{aligned}$$

whereas that of the second summand, for $z \in E$ such that $\Xi(\{z\}) = 1$, corresponds to

$$\left| \gamma \left(\xi^z + \sum_{w \in N_z \setminus \{z\}} \delta_w + \delta_z \right) - \gamma(\xi^z + \delta_z) - \gamma \left(\xi^z + \sum_{w \in N_z \setminus \{z\}} \delta_w \right) + \gamma(\xi^z) \right|,$$

which may be bounded by $\xi(N_z \setminus \{z\}) \Delta_2 \gamma$. Lemma 2.5.6 (ii) gives $\Delta_2 \gamma \leq 1$. Upper bounds for the first and second summand are then given by

$$\int_E \mathbb{E} \xi(N_z) \lambda(dz) \quad \text{and} \quad \mathbb{E} \left[\int_E \Xi(N_z \setminus \{z\}) \Xi(dz) \right],$$

respectively, where we additionally used the Fubini-Tonelli theorem for the first

summand. For the third summand, (2.29) gives

$$\begin{aligned} & \left| \mathbb{E} \int_E [\gamma(\Xi^z + \delta_z) - \gamma(\Xi^z)] \{ \Xi(dz) - \lambda(dz) \} \right| \\ &= \left| \mathbb{E} \int_E [\gamma(\Xi^z + \delta_z) - \gamma(\Xi^z)] \{ g(z, \Xi^z) - \mu(z) \} \nu(dz) \right| \\ &\leq \Delta_1 \gamma \int_E \mathbb{E} |g(z, \Xi^z) - \mu(z)| \nu(dz), \end{aligned}$$

where $\Delta_1 \gamma \leq 1$ due to Lemma 2.5.6 (i). \square

Suppose we want to approximate the law of a point process Ξ with mean measure λ by that of a Poisson process with, say, mean measure $\tilde{\lambda}$, different (but not too different) from λ . We then simply do the approximation in two steps and use the triangle inequality:

$$d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\tilde{\lambda})) \leq d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\lambda)) + d_{TV}(\text{PRM}(\lambda), \text{PRM}(\tilde{\lambda})).$$

The following proposition gives an estimate of the additional error term.

Proposition 2.5.10. *Let λ and $\tilde{\lambda}$ be two finite measures on E . Then*

$$d_{TV}(\text{PRM}(\lambda), \text{PRM}(\tilde{\lambda})) \leq \int_E |\lambda - \tilde{\lambda}|(dz)$$

Proof. Let $\Xi := \Xi_{\tilde{\lambda}} \sim \text{PRM}(\tilde{\lambda})$ and let $\Xi_{\lambda} \sim \text{PRM}(\lambda)$. Let \mathcal{A} be the generator of an immigration-death process with immigration intensity λ , unit per-capita death rate, and equilibrium distribution $\mathcal{L}(\Xi_{\lambda})$. By (2.35), $|P(\Xi \in R) - P(\Xi_{\lambda} \in R)|$ equals

$$|\mathbb{E}(\mathcal{A}\gamma)(\Xi)| = \left| \mathbb{E} \int_E [\gamma(\Xi + \delta_z) - \gamma(\Xi)] \lambda(dz) + \mathbb{E} \int_E [\gamma(\Xi - \delta_z) - \gamma(\Xi)] \Xi(dz) \right|$$

and it is sufficient to determine a uniform bound for $|\mathbb{E}(\mathcal{A}\gamma)(\Xi)|$. By (2.24), we have that for a bounded measurable function γ ,

$$\mathbb{E} \int_E [\gamma(\Xi - \delta_z) - \gamma(\Xi)] \Xi(dz) = \mathbb{E} \int_E [\gamma(\Xi_z - \delta_z) - \gamma(\Xi_z)] \tilde{\lambda}(dz),$$

where Ξ_z is the Palm process for Ξ at z . It follows from (2.25) that Ξ_z is a Poisson process with mean measure $\tilde{\lambda}$ with the addition of a deterministic atom

at z . Likewise, $\mathcal{L}(\Xi_z - \delta_z) = \mathcal{L}(\Xi)$. The integral on the right-hand side then equals

$$\mathbb{E} \int_E [\gamma(\Xi) - \gamma(\Xi + \delta_z)] \tilde{\lambda}(dz) = - \int_E \mathbb{E} [\gamma(\Xi + \delta_z) - \gamma(\Xi)] \tilde{\lambda}(dz).$$

Hence,

$$\begin{aligned} |\mathbb{E}(\mathcal{A}\gamma)(\Xi)| &= \left| \int_E \mathbb{E} [\gamma(\Xi + \delta_z) - \gamma(\Xi)] (\lambda - \tilde{\lambda})(dz) \right| \\ &\leq \int_E \Delta_1 \gamma |\lambda - \tilde{\lambda}|(dz) \leq \int_E |\lambda - \tilde{\lambda}|(dz), \end{aligned}$$

where we used Lemma 2.5.6 (i) for the last inequality. \square

The following two corollaries exemplify the use of Theorem 2.5.7. For both corollaries, we suppose that the state space E is a finite index set, called Γ , and that the point process Ξ on Γ is of the form $\sum_{\alpha \in \Gamma} I_\alpha \delta_\alpha$, with the I_α 's being indicator variables defined on the probability space (Ω, \mathcal{F}, P) . Note that Ξ is a finite point process since Γ is finite. It is also simple, since, for all $\omega \in \Omega$ and for all $\beta \in \Gamma$,

$$\Xi(\omega, \{\beta\}) = \sum_{\alpha \in \Gamma} I_\alpha(\omega) \delta_\alpha(\{\beta\}) \leq \delta_\beta(\{\beta\}) = 1.$$

Corollary 2.5.11 treats the case of independent indicator variables so the point process has no dependence whatsoever between “regions” of Γ . It gives a process analogue to Theorem 2.4.4 for Poisson approximation of sums of independent indicator variables. Likewise, Corollary 2.5.12 gives a process version of Theorem 2.4.6, where we have local dependence between the indicator variables.

Corollary 2.5.11. *Let Γ be a finite index set. Let I_α , $\alpha \in \Gamma$, be independent Bernoulli random variables with success probability $P(I_\alpha = 1) = p_\alpha \in (0, 1)$, for all $\alpha \in \Gamma$. Let $\Xi = \sum_{\alpha \in \Gamma} I_\alpha \delta_\alpha$ be a point process on Γ with intensity measure $\lambda = \sum_{\alpha \in \Gamma} p_\alpha \delta_\alpha$. Then*

$$d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\lambda)) \leq \sum_{\alpha \in \Gamma} p_\alpha^2.$$

Proof. As the I_α 's are independent, we choose neighbourhoods $N_\alpha = \{\alpha\}$, for all $\alpha \in \Gamma$. Clearly, the second and third error terms from Theorem 2.5.7 vanish, and we obtain

$$d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\boldsymbol{\lambda})) \leq \sum_{\alpha \in \Gamma} \mathbb{E}\Xi(\{\alpha\})\boldsymbol{\lambda}(\{\alpha\}) = \sum_{\alpha \in \Gamma} \mathbb{E}\Xi(\{\alpha\})^2 = \sum_{\alpha \in \Gamma} p_\alpha^2.$$

□

Corollary 2.5.12. *Let Γ be a finite index set. Let I_α , $\alpha \in \Gamma$, be Bernoulli random variables with success probability $P(I_\alpha = 1) = p_\alpha \in (0, 1)$, for all $\alpha \in \Gamma$. Let $\Xi = \sum_{\alpha \in \Gamma} I_\alpha \delta_\alpha$ be a point process on Γ with intensity measure $\boldsymbol{\lambda} = \sum_{\alpha \in \Gamma} p_\alpha \delta_\alpha$. For any choice of index $\alpha \in \Gamma$, define $\Gamma_\alpha^s \subseteq \Gamma \setminus \{\alpha\}$ to be the set of indices containing all $\beta \neq \alpha$ for which I_β is strongly dependent on I_α , and define Γ_α^w similarly as the set of indices β for which I_β is weakly dependent on I_α . Furthermore, let $Z_\alpha = \sum_{\beta \in \Gamma_\alpha^s} I_\beta$. Then,*

$$d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\boldsymbol{\lambda})) \leq \sum_{\alpha \in \Gamma} \{p_\alpha^2 + p_\alpha \mathbb{E}Z_\alpha + \mathbb{E}I_\alpha Z_\alpha + \eta_\alpha\},$$

where

$$\eta_\alpha = \mathbb{E}|\mathbb{E}\{(I_\beta, \beta \in \Gamma_\alpha^w)\} - p_\alpha|.$$

Proof. For each $\alpha \in \Gamma$, we choose the neighbourhood $N_\alpha = \Gamma_\alpha^s \cup \{\alpha\}$. Let $\Xi^\alpha = \sum_{\beta \in \Gamma_\alpha^w} I_\beta \delta_\beta$ be the configuration of Ξ outside N_α . From the proof of Theorem 2.5.7,

$$\begin{aligned} d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\boldsymbol{\lambda})) &\leq \sum_{\alpha \in \Gamma} \mathbb{E}\Xi(N_\alpha)\boldsymbol{\lambda}(\{\alpha\}) + \sum_{\alpha \in \Gamma} \mathbb{E}[\Xi(N_\alpha \setminus \{\alpha\})\Xi(\{\alpha\})] \\ &\quad + \left| \sum_{\alpha \in \Gamma} \mathbb{E}[\gamma(\Xi^\alpha + \delta_\alpha) - \gamma(\Xi^\alpha)] \{\Xi(\{\alpha\}) - \boldsymbol{\lambda}(\{\alpha\})\} \right|. \end{aligned}$$

The first of these three error terms equals

$$\begin{aligned} \sum_{\alpha \in \Gamma} \mathbb{E}\Xi(\Gamma_\alpha^s \cup \{\alpha\})p_\alpha &= \sum_{\alpha \in \Gamma} p_\alpha \left(\mathbb{E}\Xi(\{\alpha\}) + \sum_{\beta \in \Gamma_\alpha^s} \mathbb{E}\Xi(\{\beta\}) \right) \\ &= \sum_{\alpha \in \Gamma} \{p_\alpha^2 + p_\alpha \mathbb{E}Z_\alpha\}, \end{aligned}$$

and the second error term equals

$$\sum_{\alpha \in \Gamma} \mathbb{E}[\Xi(\Gamma_{\alpha}^s) \Xi(\{\alpha\})] = \sum_{\alpha \in \Gamma} \mathbb{E} \left[\left(\sum_{\beta \in \Gamma_{\alpha}^s} I_{\beta} \right) I_{\alpha} \right] = \sum_{\alpha \in \Gamma} \mathbb{E}[I_{\alpha} Z_{\alpha}].$$

It remains to observe that the third error term is bounded by

$$\sum_{\alpha \in \Gamma} |\mathbb{E}(I_{\alpha} - p_{\alpha})[\gamma(\Xi^{\alpha} + \delta_{\alpha}) - \gamma(\Xi^{\alpha})]| \leq \sum_{\alpha \in \Gamma} \eta_{\alpha} \Delta_1 \gamma,$$

and $\Delta_1 \gamma \leq 1$ by Lemma 2.5.6 (i). □

When comparing Theorems 2.4.4 and 2.4.6 (which treated the concrete example of $\Gamma = \{1, \dots, n\}$) with their process analogues, Corollaries 2.5.11 and 2.5.12, we see that the respective error bounds are of the same form except for the multiplicative factors in λ that are absent in the process results. As these factors decrease towards zero with increasing λ , the lack of them in the process results shows that Poisson process approximation gives bigger errors than Poisson approximation. The reason for this is that the total variation distance is so strong that it does not allow for even the smallest shifts in the positions of points on the carrier space. That is, if the sets of placements of the points of two point processes in a carrier space are disjoint, then, even if the points of the two processes are placed close to each other with respect to some metric on the carrier space, the total variation distance takes value 1, the maximum value it can take for a pair of probability distributions. As a consequence, the total variation distance is not at all suited for approximating a process on a lattice in \mathbb{R}^d by a process with a continuous intensity over \mathbb{R}^d . An example was given in Chapter 1.

The hope is to find a way to recover multiplicative factors that decrease with increasing λ when approximating a point process by a Poisson process. One way to do this is to use a metric that is weaker than the total variation metric and able to exploit the closeness in the positions of the points of the two processes. We would thus compare $\mathbb{E}h(\Xi_n)$ by $\mathbb{E}h(\Xi)$ for a set of functionals h that is smaller than the one used for approximation in total variation, and whose elements are not too sensitive to small differences in the positions of points. Barbour and Brown (1992) and Barbour et al. (1992) constructed a suitable weaker metric, the d_2 -metric. We will give their results in Section 2.6. Another way to improve Poisson process approximation is to consider marked point

processes (in situations where the use of marks makes sense). As we see in the following section, it is sometimes possible to actually recover the sharper results from Poisson approximation. Consider, for instance, the basic case where we associate a point z from a carrier space E to each indicator variable I_i , when we know that the law of $W = \sum_{i=1}^n I_i$ is close to the Poisson distribution with parameter $\mathbb{E}W$. Then, fixing any z , the process $\sum_{i=1}^n I_i \delta_z = W \delta_z$ gives W points at position z and it is clear that Poisson process approximation is the same as Poisson approximation for W .

2.5.5 Approximation of point processes with i.i.d. marks – Michel's argument

Suppose Ξ is a marked point process of the form $\sum_{i=1}^n I_i \delta_{X_i}$, where the Bernoulli random variables I_1, \dots, I_n are independent of the i.i.d. E -valued marks X_1, \dots, X_n . We may then use an argument made by Michel (1988) to show that the total variation distance between $\mathcal{L}(\Xi)$ and the law of a Poisson process with mean measure $\mathbb{E}\Xi$ is smaller than or equal to the total variation distance between the law of W and that of a Poisson random variable with mean $\mathbb{E}W$. We may thus use Theorem 2.4.4 to estimate the approximation error between $\mathcal{L}(\Xi)$ and $\text{PRM}(\mathbb{E}\Xi)$.

Theorem 2.5.13. *For each integer $n \geq 1$, let I_1, \dots, I_n be Bernoulli random variables with probability of success $P(I_i = 1) = p_i \in (0, 1)$. Let E be a locally compact separable metric space and let X_1, \dots, X_n be i.i.d. E -valued random variables, independent of the I_i 's. Moreover, let $\Xi = \sum_{i=1}^n I_i \delta_{X_i}$ and let $W = \sum_{i=1}^n I_i$. Then,*

$$d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\mathbb{E}\Xi)) \leq d_{TV}(\mathcal{L}(W), \text{Poi}(\mathbb{E}W)).$$

Proof. Let Z_1, \dots, Z_n be i.i.d. random variables with distribution $\mathcal{L}(X_1)$, and let them be independent of W . Then the process $\sum_{j=1}^W \delta_{Z_j}$ has the same distribution as the process of interest Ξ . Furthermore, note that a $\text{PRM}(\mathbb{E}\Xi)$ can be realised as $\sum_{j=1}^{W^*} \delta_{Z_j}$, where $W^* \sim \text{Poi}(\mathbb{E}W)$ is independent of the Z_j 's. Then,

using (2.6) for the total variation distance,

$$\begin{aligned}
& d_{TV}(\mathcal{L}(\Xi), \text{PRM}(\mathbb{E}\Xi)) \\
&= \sup_R \{P(\Xi \in R) - P(\text{PRM}(\mathbb{E}\Xi) \in R)\} \\
&= \sup_R \left\{ P\left(\sum_{j=1}^W \delta_{Z_j} \in R\right) - P\left(\sum_{j=1}^{W^*} \delta_{Z_j} \in R\right) \right\} \\
&= \sup_R \left\{ \sum_{l=0}^n P\left(\sum_{j=1}^W \delta_{Z_j} \in R, W = l\right) - \sum_{l=0}^{\infty} P\left(\sum_{j=1}^{W^*} \delta_{Z_j} \in R, W^* = l\right) \right\} \\
&= \sup_R \left\{ \sum_{l=0}^n P\left(\sum_{j=1}^l \delta_{Z_j} \in R\right) P(W = l) \right. \\
&\quad \left. - \sum_{l=0}^{\infty} P\left(\sum_{j=1}^l \delta_{Z_j} \in R\right) P(W^* = l) \right\} \\
&\leq \sup_R \sum_{l=0}^n P\left(\sum_{j=1}^l \delta_{Z_j} \in R\right) \{P(W = l) - P(W^* = l)\}_+ \\
&\leq \sum_{l=0}^n \{P(W = l) - P(W^* = l)\}_+,
\end{aligned}$$

where $\{\cdot\}_+ = \max(\cdot, 0)$. Now define

$$B_0 = \{l \in \{1, \dots, n\} : P(W = l) > P(W^* = l)\}.$$

Then

$$\begin{aligned}
\sum_{l=0}^n \{P(W = l) - P(W^* = l)\}_+ &= \sum_{l \in B_0} \{P(W = l) - P(W^* = l)\} \\
&= P(W \in B_0) - P(W^* \in B_0) \\
&\leq \sup_{B \subseteq \mathbb{Z}_+} |P(W \in B) - P(W^* \in B)| \\
&= d_{TV}(\mathcal{L}(W), \text{Poi}(\mathbb{E}W)).
\end{aligned}$$

□

2.6 Improved rates for Poisson process approximation using the d_2 -distance

This section gives the results by Barbour and Brown (1992) and Barbour et al. (1992) for Poisson process approximation in a metric that is weaker than the total variation metric. As before in Section 2.5, we assume that E is a locally compact separable metric space. Let d_0 be a metric on E that is bounded by 1. We now define metrics on both the space $M_p(E)$ of finite point measures over E and on the set of probability measures over $M_p(E)$. Let \mathcal{K} denote the set of functions $\kappa : E \rightarrow \mathbb{R}$ such that

$$s_1(\kappa) = \sup_{z_1 \neq z_2 \in E} \frac{|\kappa(z_1) - \kappa(z_2)|}{d_0(z_1, z_2)} < \infty,$$

which implies that for all $z_1 \neq z_2 \in E$, $|\kappa(z_1) - \kappa(z_2)| \leq s_1(\kappa)d_0(z_1, z_2)$. Thus each function $\kappa \in \mathcal{K}$ is Lipschitz continuous with constant $s_1(\kappa)$. Define a distance d_1 between two finite measures ρ and σ over E by

$$d_1(\rho, \sigma) = \begin{cases} 1, & \text{if } \rho(E) \neq \sigma(E), \\ \frac{1}{\rho(E)} \sup_{\kappa \in \mathcal{K}} \frac{|\int_E \kappa d\rho - \int_E \kappa d\sigma|}{s_1(\kappa)}, & \text{if } \rho(E) = \sigma(E). \end{cases} \quad (2.37)$$

Note that d_1 is bounded by 1. We can use d_1 as distance between point measures in $M_p(E)$. The d_1 -distance is then a Wasserstein metric induced by d_0 over point measures on E . Suppose that we have two point configurations $\xi_1, \xi_2 \in M_p(E)$ with the same number of points $|\xi_1| = |\xi_2| = m$. It then follows from (2.9) that $d_1(\xi_1, \xi_2)$ can be interpreted as the average distance between the points (z_{11}, \dots, z_{1m}) and (z_{21}, \dots, z_{2m}) of ξ_1 and ξ_2 under their closest matching, i.e.

$$d_1(\xi_1, \xi_2) = \min_{\pi \in S_m} \frac{1}{m} \sum_{j=1}^m d_0(z_{1j}, z_{2\pi(j)}). \quad (2.38)$$

See Figure 2.4 for an illustration.

We establish a useful result for the d_1 -distance between two point configurations differing only in one point, i.e. for $d_1(\xi + \delta_z, \xi + \delta_w)$, where $\xi \in M_p(E)$ and $z \neq w \in E$. To achieve this, note that

$$\begin{aligned} \int_E \kappa(v)(\xi + \delta_z)(dv) - \int_E \kappa(v)(\xi + \delta_w)(dv) &= \int_E \kappa(v)\delta_z(dv) - \int_E \kappa(v)\delta_w(dv) \\ &= \kappa(z) - \kappa(w), \end{aligned}$$

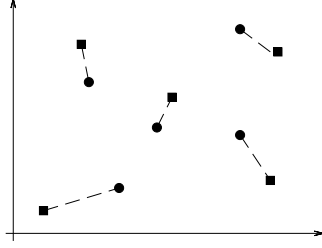


Figure 2.4: Two point configurations ξ_1 (bullets) and ξ_2 (squares) with five points each on $E = \mathbb{R}_+^2$. The dashed lines represent the d_0 -distances (here the Euclidean distances bounded by 1) between the closest matchings.

and we therefore obtain, using $s_1(\kappa)^{-1} \leq d_0(z, w)/|\kappa(z) - \kappa(w)|$ from Lipschitz continuity of $\kappa \in \mathcal{K}$, that

$$\begin{aligned} & d_1(\xi + \delta_z, \xi + \delta_w) \\ &= \frac{1}{|\xi| + 1} \sup_{\kappa \in \mathcal{K}} \frac{d_0(z, w) \left| \int_E \kappa(v)(\xi + \delta_z)(dv) - \int_E \kappa(v)(\xi + \delta_w)(dv) \right|}{|\kappa(z) - \kappa(w)|} \\ &= \frac{1}{|\xi| + 1} d_0(z, w). \end{aligned} \quad (2.39)$$

We next construct a metric d_2 that is a Wasserstein metric induced by d_1 over probability measures on $M_p(E)$. Let \mathcal{H} denote the set of functions $h : M_p(E) \rightarrow \mathbb{R}$ such that

$$s_2(h) = \sup_{\xi_1 \neq \xi_2 \in M_p(E)} \frac{|h(\xi_1) - h(\xi_2)|}{d_1(\xi_1, \xi_2)} < \infty, \quad (2.40)$$

i.e. each function $h \in \mathcal{H}$ is Lipschitz continuous with constant $s_2(h)$. We define a distance d_2 between probability measures μ and ν over $M_p(E)$ by

$$d_2(\mu, \nu) = \frac{1}{s_2(h)} \sup_{h \in \mathcal{H}} \left| \int_{M_p(E)} h d\mu - \int_{M_p(E)} h d\nu \right|. \quad (2.41)$$

Note that d_2 is bounded by 1. By setting $\tilde{h} := h/s_2(h)$ for each $h \in \mathcal{H}$, we may

equivalently write

$$d_2(\mu, \nu) = \sup_{\tilde{h} \in \tilde{\mathcal{H}}} \left| \int_{M_p(E)} \tilde{h} d\mu - \int_{M_p(E)} \tilde{h} d\nu \right|,$$

where $\tilde{\mathcal{H}} = \{\tilde{h} : M_p(E) \rightarrow \mathbb{R}; |\tilde{h}(\xi_1) - \tilde{h}(\xi_2)| \leq d_1(\xi_1, \xi_2) \leq 1, \forall \xi_1, \xi_2 \in M_p(E)\}$. By (2.7), the test functions \tilde{h} used to define the total variation distance between two probability measures μ and ν on $M_p(E)$ satisfy $|\tilde{h}(\xi_1) - \tilde{h}(\xi_2)| \leq 1$, for any $\xi_1, \xi_2 \in M_p(E)$. The set of test functions used to define d_2 is thus contained in the set of test functions used to define the total variation distance. It follows that for any probability measures μ and ν on $M_p(E)$,

$$d_2(\mu, \nu) \leq d_{TV}(\mu, \nu). \quad (2.42)$$

For a point process Ξ on E with intensity measure λ , let \mathcal{A} be the generator of an immigration-death process with immigration intensity λ , unit per-capita death rate, and equilibrium distribution $\text{PRM}(\lambda)$, i.e. let \mathcal{A} be as defined in (2.30). Moreover, let γ be as defined in Proposition 2.5.4 for $h \in \mathcal{H}$. Then, by Proposition 2.5.5,

$$\begin{aligned} |\mathbb{E}(\mathcal{A}\gamma)(\Xi)| &= |\mathbb{E}h(\Xi) - \text{PRM}(\lambda)(h)| \\ &= \left| \int_{M_p(E)} h d\mathcal{L}(\Xi) - \int_{M_p(E)} h d\text{PRM}(\lambda) \right|, \end{aligned} \quad (2.43)$$

By (2.41), finding an upper bound on $d_2(\mathcal{L}(\Xi), \text{PRM}(\lambda))$ is thus equivalent to finding a uniform upper bound on $|\mathbb{E}(\mathcal{A}\gamma)(\Xi)|/s_2(h)$. For the latter we need smoothness estimates of the solution γ of the Stein equation when $h \in \mathcal{H}$ (instead of h being an indicator function as in Section 2.5). We determine such estimates in Lemmas 2.6.3 and 2.6.4 below. In order to prove these, we first need the following two lemmas. Note that our proof of Lemma 2.6.2 below corrects a slight mistake in Barbour and Brown (1992) and Barbour et al. (1992).

Lemma 2.6.1. *Let Z and Z^0 be immigration-death processes on E with immigration intensity λ and unit per-capita death rate, where $\lambda = \lambda(E) < \infty$, Z has point configuration $\xi = \sum_{j=1}^{|\xi|} \delta_{z_j} \in M_p(E)$ at time $t = 0$, and Z^0 has no initial particles. Let $T_1, T_2, \dots, T_{|\xi|}$ be independent $\text{Exp}(1)$ -random variables, independent of Z^0 . Then*

$$Z_t \stackrel{d}{=} Z_t^0 + D_t, \quad \text{for all } t \in \mathbb{R}_+,$$

where $D_t = \sum_{j=1}^{\lfloor \xi \rfloor} \delta_{z_j} I_{\{T_j > t\}}$ is a pure death process and $Z_t^0 \sim \text{PRM}((1-e^{-t})\lambda)$.

Proof. See Proposition 3.5 in Xia (2005). \square

Lemma 2.6.2. *Let $Z = \{Z_t, t \in \mathbb{R}_+\}$ be an immigration-death process on \mathbb{Z}_+ with constant immigration rate $\lambda > 0$ and unit per-capita death rate, with k initial particles, i.e. $P(Z_0 = k) = 1$. Then*

$$\int_0^\infty e^{-t} \mathbb{E}[(Z_t + 1)^{-1}] dt \leq \left(\frac{1}{\lambda} + \frac{1}{k+1} \right) (1 - e^{-\lambda}).$$

Proof. The particles that are alive in the population at time t can be grouped into two categories: those among the k particles that were in the population from time 0, and those that arrived later. By Lemma 2.6.1, we may thus express the number Z_t of particles in the population at time t as the sum of independent random variables $X_t \sim \text{Bin}(k, e^{-t})$ and $Y_t \sim \text{Poi}(\lambda(1 - e^{-t}))$. We then have

$$\mathbb{E}[(Z_t + 1)^{-1}] \leq \mathbb{E}[(X_t + 1)^{-1}] \quad \text{and} \quad \mathbb{E}[(Z_t + 1)^{-1}] \leq \mathbb{E}[(Y_t + 1)^{-1}].$$

On the one hand, setting $m = l + 1$,

$$\begin{aligned} \mathbb{E}[(X_t + 1)^{-1}] &= \sum_{l=0}^k \frac{1}{l+1} \binom{k}{l} (e^{-t})^l (1 - e^{-t})^{k-l} \\ &= (1 - e^{-t})^k \left(\frac{e^{-t}}{1 - e^{-t}} \right)^{-1} \sum_{m=1}^{k+1} \frac{k!}{m!(k-m+1)!} \left(\frac{e^{-t}}{1 - e^{-t}} \right)^m \\ &= (1 - e^{-t})^{k+1} \frac{e^t}{k+1} \left\{ \sum_{m=0}^{k+1} \frac{(k+1)!}{m!(k+1-m)!} \left(\frac{e^{-t}}{1 - e^{-t}} \right)^m - 1 \right\} \\ &= (1 - e^{-t})^{k+1} \frac{e^t}{k+1} \left\{ \left(1 + \frac{e^{-t}}{1 - e^{-t}} \right)^{k+1} - 1 \right\} \\ &= \frac{e^t}{k+1} \{ 1 - (1 - e^{-t})^{k+1} \}. \end{aligned}$$

On the other hand, setting $m = l + 1$,

$$\mathbb{E}[(Y_t + 1)^{-1}] = e^{-\lambda t} \sum_{l=0}^{\infty} \frac{\lambda_t^l}{(l+1)!} = \frac{e^{-\lambda t}}{\lambda t} \left(\sum_{m=0}^{\infty} \frac{\lambda_t^m}{m!} - 1 \right) = \frac{1 - e^{-\lambda(1 - e^{-t})}}{\lambda(1 - e^{-t})}.$$

With τ such that $e^{-\tau} = \lambda/(\lambda + k + 1)$, we now have

$$\begin{aligned}
 & \int_0^\infty e^{-t} \mathbb{E}[(Z_t + 1)^{-1}] dt \\
 & \leq \int_0^\tau e^{-t} \mathbb{E}[(X_t + 1)^{-1}] dt + \int_\tau^\infty e^{-t} \mathbb{E}[(Y_t + 1)^{-1}] dt \\
 & \leq \int_0^\tau \frac{1}{k+1} \{1 - (1 - e^{-t})^{k+1}\} dt + \frac{1 - e^{-\lambda}}{\lambda} \int_\tau^\infty \frac{e^{-t}}{1 - e^{-t}} dt.
 \end{aligned} \tag{2.44}$$

The second of these integrals equals

$$-\frac{1 - e^{-\lambda}}{\lambda} \log(1 - e^{-\tau}) = \frac{1 - e^{-\lambda}}{\lambda} \log\left(1 + \frac{\lambda}{k+1}\right) \leq \frac{1 - e^{-\lambda}}{k+1}, \tag{2.45}$$

since $-\log(1 - z) \leq z/(1 - z)$ for $z = \lambda/(\lambda + k + 1) < 1$. Furthermore, due to

$$1 - (1 - e^{-t})^{k+1} = e^{-t} \cdot \frac{1 - (1 - e^{-t})^{k+1}}{1 - (1 - e^{-t})} = e^{-t} \sum_{j=0}^k (1 - e^{-t})^j,$$

the first integral equals

$$\frac{1}{k+1} \sum_{j=0}^k \int_0^\tau e^{-t} (1 - e^{-t})^j dt = \frac{1}{k+1} \sum_{j=1}^{k+1} \frac{(1 - e^{-\tau})^j}{j}. \tag{2.46}$$

By setting $x := \lambda/(k+1)$, we may rewrite the expression in (2.46) as $(x/\lambda) \cdot \sum_{j=1}^{k+1} [j(1+x)^j]^{-1}$, which is smaller than

$$\begin{aligned}
 \frac{x}{\lambda} \sum_{j=1}^{k+1} \frac{1}{(1+x)^j} &= \frac{x}{\lambda} \left[\sum_{j=0}^\infty \frac{1}{(1+x)^j} - 1 - \sum_{j=k+2}^\infty \frac{1}{(1+x)^j} \right] \\
 &= \frac{x}{\lambda} \left[\frac{1}{x} - \frac{(1+x)^{-k-2}}{1 - (1+x)^{-1}} \right] = \frac{1}{\lambda} \left[1 - \frac{1}{(1+x)^{k+1}} \right] \\
 &= \frac{1}{\lambda} \left(1 - e^{-\frac{\lambda \log(1+x)}{x}} \right) \leq \frac{1 - e^{-\lambda}}{\lambda},
 \end{aligned} \tag{2.47}$$

where we used $\log(1+x) \leq x$ for the last inequality. By combining (2.44), (2.45) and (2.47), we obtain the lemma. \square

The following two lemmas give smoothness estimates of the solution γ of the Stein equation when $h \in \mathcal{H}$. They are the counterparts of (i) and (ii) of Lemma 2.5.6, respectively, for the smaller class of Lipschitz continuous functions \mathcal{H} .

Lemma 2.6.3. *Let λ be a finite measure over E with $\lambda(E) = \lambda$. Let $\gamma : M_p(E) \rightarrow \mathbb{R}$ be defined as in Proposition 2.5.4, where $h : M_p(E) \rightarrow \mathbb{R}$ is any function in \mathcal{H} . Then, for any $\xi \in M_p(E)$,*

$$\Delta_1 \gamma \leq s_2(h) \left(1 \wedge \frac{1.65}{\sqrt{\lambda}} \right).$$

Proof. From Proposition 2.5.4, we have that for any $\xi = \sum_{j \in J} \delta_{w_j} \in M_p(E)$, where $J \subseteq \mathbb{N}$, and for any $z \in E$,

$$\gamma(\xi + \delta_z) - \gamma(\xi) = \int_0^\infty \{ \mathbb{E}^\xi h(Z_t) - \mathbb{E}^{\xi + \delta_z} h(Z_t) \} dt,$$

where $Z = \{Z_t, t \in \mathbb{R}_+\}$ is the immigration-death process on E with immigration intensity λ and unit per-capita death rate. Let Z be realised under \mathbb{P}^ξ and let T be an exponential random variable with parameter 1, independent of Z . Then the process Z' defined by $Z'_t = Z_t + \delta_z I_{\{T > t\}}$ has distribution $\mathbb{P}^{\xi + \delta_z}$. Moreover, let Z^0 be realised under \mathbb{P}^0 and let D be a pure death process with unit per-capita death rate starting with $D_0 = \xi$. Then, $Z_t = Z_t^0 + D_t$ by Lemma 2.6.1, and

$$\begin{aligned} & \gamma(\xi + \delta_z) - \gamma(\xi) \\ &= \int_0^\infty e^{-t} \mathbb{E}^\xi [h(Z_t^0 + D_t) - h(Z_t^0 + D_t + \delta_z)] dt \\ &= \int_0^\infty e^{-t} \sum_{\eta \in \mathcal{N}} \mathbb{E} [h(Z_t^0 + \eta) - h(Z_t^0 + \eta + \delta_z)] P(D_t = \eta) dt, \end{aligned} \tag{2.48}$$

where $\mathcal{N} = \{\sum_{j \in J'} \delta_{w_j} ; J' \subseteq J\}$. We first show that $\Delta_1 \gamma \leq s_2(h)$. Using Lipschitz continuity of h , as well as the fact that the d_1 -distance between point configurations of different sizes is 1 (see (2.40) and (2.37), respectively), we find that

$$|\gamma(\xi + \delta_z) - \gamma(\xi)| \leq \int_0^\infty e^{-t} \sum_{\eta \in \mathcal{N}} P(D_t = \eta) dt \leq s_2(h) \int_0^\infty e^{-t} dt = s_2(h).$$

In order to show that $\Delta_1 \gamma \leq s_2(h)(1.65/\sqrt{\lambda})$, note first that

$$\begin{aligned} & \mathbb{E} [h(Z_t^0 + \eta) - h(Z_t^0 + \eta + \delta_z)] \\ &= \sum_{k \geq 0} P(|Z_t^0| = k) \mathbb{E} [h(Z_t^0 + \eta) - h(Z_t^0 + \eta + \delta_z) \mid |Z_t^0| = k] \\ &= P(|Z_t^0| = 0) h(\eta) + \sum_{k \geq 0} \{P(|Z_t^0| = k+1) \mathbb{E} [h(Z_t^0 + \eta) \mid |Z_t^0| = k+1] \\ &\quad - P(|Z_t^0| = k) \mathbb{E} [h(Z_t^0 + \eta + \delta_z) \mid |Z_t^0| = k]\}. \end{aligned}$$

For the part in curly brackets we use $|a_1 a_2 - b_1 b_2| \leq a_2 |a_1 - b_1| + b_1 |a_2 - b_2|$, where

$$\begin{aligned} a_1 &:= P(|Z_t^0| = k+1), & a_2 &:= \mathbb{E} [h(Z_t^0 + \eta) \mid |Z_t^0| = k+1], \\ b_1 &:= P(|Z_t^0| = k), & b_2 &:= \mathbb{E} [h(Z_t^0 + \eta + \delta_z) \mid |Z_t^0| = k]. \end{aligned}$$

We have

$$\begin{aligned} |a_2 - b_2| &= \frac{1}{\lambda} \left| \int_E \mathbb{E} [h(Z_t^0 + \eta + \delta_w) - h(Z_t^0 + \eta + \delta_z) \mid |Z_t^0| = k] \lambda(dw) \right| \\ &\leq \frac{1}{\lambda} \int_E \mathbb{E} [s_2(h) d_1(Z_t^0 + \eta + \delta_w, Z_t^0 + \eta + \delta_z) \mid |Z_t^0| = k] \lambda(dw) \\ &\leq \frac{s_2(h)}{\lambda} \int_E \mathbb{E} [(|Z_t^0| + |\eta| + 1)^{-1} d_0(w, z) \mid |Z_t^0| = k] \lambda(dw) \\ &\leq s_2(h)(k + |\eta| + 1)^{-1}, \end{aligned} \tag{2.49}$$

where we used Lipschitz continuity of h , (2.39), and boundedness of d_0 by 1 for the first, second and third inequalities, respectively. It follows that

$$\sum_{k \geq 0} b_1 |a_2 - b_2| \leq s_2(h) \sum_{k \geq 0} \frac{P(|Z_t^0| = k)}{k + |\eta| + 1} \leq s_2(h) \mathbb{E} [(|Z_t^0| + 1)^{-1}].$$

Furthermore, note that the use of the function $h - (\inf_{\xi} h + \sup_{\xi} h)/2$ instead of h leaves (2.48) unchanged. Therefore, we may use $\sup_{\xi} |h| = s_2(h)/2$, which entails the following two bounds:

$$\begin{aligned} a_2 &\leq |\mathbb{E} [h(Z_t^0 + \eta) \mid |Z_t^0| = k]| \leq \frac{s_2(h)}{2}, \\ P(|Z_t^0| = 0) h(\eta) &\leq \frac{s_2(h)}{2} P(|Z_t^0| = 0). \end{aligned} \tag{2.50}$$

We obtain

$$\begin{aligned} & \left| \mathbb{E} \left[h(Z_t^0 + \eta) - h(Z_t^0 + \eta + \delta_z) \right] \right| \\ & \leq s_2(h) \left\{ \frac{1}{2} P(|Z_t^0| = 0) + \frac{1}{2} \sum_{k \geq 0} |P(|Z_t^0| = k+1) - P(|Z_t^0| = k)| \right\} \\ & \quad + s_2(h) \mathbb{E} [(|Z_t^0| + 1)^{-1}]. \end{aligned} \quad (2.51)$$

By Lemma 2.6.1, we have $|Z_t^0| \sim \text{Poi}(\lambda_t)$ with $\lambda_t := \lambda(1 - e^{-t})$. Thus,

$$\mathbb{E} [(|Z_t^0| + 1)^{-1}] = \frac{e^{-\lambda_t}}{\lambda_t} \sum_{k \geq 0} \frac{\lambda_t^{k+1}}{(k+1)!} = \frac{1 - e^{-\lambda_t}}{\lambda_t}. \quad (2.52)$$

Furthermore, note that

$$P(|Z_t^0| = k+1) - P(|Z_t^0| = k) = \frac{e^{-\lambda_t} \lambda_t^k}{k!} \left(\frac{\lambda_t}{k+1} - 1 \right),$$

and that, if $k < \lambda_t - 1$, then $P(|Z_t^0| = k+1) > P(|Z_t^0| = k)$, and else, if $k > \lambda_t - 1$, then $P(|Z_t^0| = k+1) < P(|Z_t^0| = k)$. Thus,

$$\begin{aligned} & \frac{1}{2} P(|Z_t^0| = 0) + \frac{1}{2} \sum_{k \geq 0} |P(|Z_t^0| = k+1) - P(|Z_t^0| = k)| \\ & = \frac{1}{2} P(|Z_t^0| = 0) + \frac{1}{2} \sum_{k=0}^{\lfloor \lambda_t - 1 \rfloor} [P(|Z_t^0| = k+1) - P(|Z_t^0| = k)] \\ & \quad + \frac{1}{2} \sum_{k=\lceil \lambda_t - 1 \rceil}^{\infty} [P(|Z_t^0| = k) - P(|Z_t^0| = k+1)] \end{aligned}$$

equals

$$\begin{aligned} & \frac{1}{2} \{ P(|Z_t^0| = \lfloor \lambda_t - 1 \rfloor + 1) + P(|Z_t^0| = \lceil \lambda_t - 1 \rceil) \} \leq \max_{k \geq 0} P(|Z_t^0| = k) \\ & \leq \frac{1}{\sqrt{2e\lambda_t}}, \end{aligned} \quad (2.53)$$

where the last inequality is due to Proposition A.2.7 in Barbour et al. (1992). In addition to the estimate (2.51) with (2.52) and (2.53), we get the following

more direct estimate from Lipschitz continuity of h :

$$|\mathbb{E}[h(Z_t^0 + \eta) - h(Z_t^0 + \eta + \delta_z)]| \leq s_2(h). \quad (2.54)$$

Choose τ such that $e^{-\tau} = 1 - \lambda^{-1}$. With (2.51)-(2.54), the following then holds for any $\eta \in \mathcal{N}$:

$$\begin{aligned} & \frac{1}{s_2(h)} \int_0^\infty e^{-t} |\mathbb{E}[h(Z_t^0 + \eta) - h(Z_t^0 + \eta + \delta_z)]| dt \\ & \leq \int_0^\tau e^{-t} dt + \int_\tau^\infty e^{-t} \left(\frac{1}{\sqrt{2e\lambda_t}} + \frac{1 - e^{-\lambda_t}}{\lambda_t} \right) dt \\ & \leq \int_0^\tau e^{-t} dt + \frac{1}{\sqrt{2e\lambda}} \int_\tau^\infty \frac{e^{-t}}{\sqrt{1 - e^{-t}}} dt + \frac{1}{\lambda} \int_\tau^\infty \frac{e^{-t}}{1 - e^{-t}} dt \\ & = 1 - e^{-\tau} + \frac{2 - 2\sqrt{1 - e^{-\tau}}}{\sqrt{2e\lambda}} - \frac{\log(1 - e^{-\tau})}{\lambda} = \frac{1}{\lambda} + \sqrt{\frac{2}{e\lambda}} - \frac{1}{\lambda} \sqrt{\frac{2}{e}} + \frac{\log \lambda}{\lambda} \\ & \leq \frac{1}{\sqrt{\lambda}} \left(\frac{0.14223 + \log \lambda}{\sqrt{\lambda}} + 0.86 \right) \leq \frac{1}{\sqrt{\lambda}} (0.79 + 0.86) = \frac{1.65}{\sqrt{\lambda}}. \end{aligned}$$

Thus,

$$\begin{aligned} \Delta_1 \gamma & \leq \sum_{\eta \in \mathcal{N}} P(D_t = \eta) \int_0^\infty e^{-t} |\mathbb{E}[h(Z_t^0 + \eta) - h(Z_t^0 + \eta + \delta_z)]| dt \\ & \leq \frac{1.65 s_2(h)}{\sqrt{\lambda}} \sum_{\eta \in \mathcal{N}} P(D_t = \eta) \leq \frac{1.65 s_2(h)}{\sqrt{\lambda}}. \end{aligned}$$

□

Lemma 2.6.4. *Under the conditions of Lemma 2.6.3,*

$$\Delta_2 \gamma \leq s_2(h) \left\{ 1 \wedge \frac{2}{\lambda} \left(1 + 2 \log_+ \left(\frac{\lambda}{2} \right) \right) \right\}.$$

Proof. As in the proof of Lemma 2.5.6 (ii), we may write, for any $\xi = \sum_{j \in J} \delta_{w_j} \in M_p(E)$, where $J \subseteq \mathbb{N}$, and for any $z, w \in E$,

$$\begin{aligned} & \gamma(\xi + \delta_z + \delta_w) - \gamma(\xi + \delta_z) - \gamma(\xi + \delta_w) + \gamma(\xi) \\ & = - \int_0^\infty \mathbb{E}^\xi [h(Z_t + \delta_z + \delta_w) - h(Z_t + \delta_z) - h(Z_t + \delta_w) + h(Z_t)] e^{-2t} dt, \end{aligned}$$

where $Z = \{Z_t, t \in \mathbb{R}_+\}$ is an immigration-death process on E with immigration intensity λ and unit per-capita death rate realised under \mathbb{P}^ξ . Let Z^0 be realised under \mathbb{P}^0 and let D be a pure death process with unit per-capita death rate starting with $D_0 = \xi$. Then $Z_t = Z_t^0 + D_t$ by Lemma 2.6.1, and

$$\begin{aligned} & \gamma(\xi + \delta_z + \delta_w) - \gamma(\xi + \delta_z) - \gamma(\xi + \delta_w) + \gamma(\xi) \\ &= - \int_0^\infty e^{-2t} \sum_{\eta \in \mathcal{N}} \mathbb{E} [h(Z_t^0 + \eta + \delta_z + \delta_w) - h(Z_t^0 + \eta + \delta_z) \\ & \quad - h(Z_t^0 + \eta + \delta_w) + h(Z_t^0 + \eta)] P(D_t = \eta) dt, \end{aligned} \tag{2.55}$$

where $\mathcal{N} = \{\sum_{j \in J'} \delta_{w_j} ; J' \subseteq J\}$. We first show that $\Delta_2 \gamma \leq s_2(h)$. For any $\eta \in \mathcal{N}$, it follows from Lipschitz continuity of h that

$$\begin{aligned} & |\mathbb{E} [h(Z_t^0 + \eta + \delta_z + \delta_w) - h(Z_t^0 + \eta + \delta_z) - h(Z_t^0 + \eta + \delta_w) + h(Z_t^0 + \eta)]| \\ & \leq \mathbb{E} |h(Z_t^0 + \eta + \delta_z + \delta_w) - h(Z_t^0 + \eta + \delta_z)| + \mathbb{E} |h(Z_t^0 + \eta + \delta_w) - h(Z_t^0 + \eta)| \\ & \leq 2s_2(h). \end{aligned}$$

Then,

$$\Delta_2 \gamma \leq 2s_2(h) \int_0^\infty e^{-2t} \sum_{\eta \in \mathcal{N}} P(D_t = \eta) dt \leq 2s_2(h) \int_0^\infty e^{-2t} dt = s_2(h).$$

For the λ -dependent bound, note that the expectation in (2.55) may be rewritten as

$$\begin{aligned} & \sum_{k \geq -1} \{ P(|Z_t^0| = k) \mathbb{E} [h(Z_t^0 + \eta + \delta_z + \delta_w) \mid |Z_t^0| = k] \\ & \quad - P(|Z_t^0| = k+1) \mathbb{E} [h(Z_t^0 + \eta + \delta_z) + h(Z_t^0 + \eta + \delta_w) \mid |Z_t^0| = k+1] \\ & \quad + P(|Z_t^0| = k+2) \mathbb{E} [h(Z_t^0 + \eta) \mid |Z_t^0| = k+2] \} \\ & + P(|Z_t^0| = 0) h(\eta). \end{aligned}$$

We add and subtract both $P(|Z_t^0| = k) \mathbb{E} [h(Z_t^0 + \eta + \delta_z) \mid |Z_t^0| = k+1]$ and $P(|Z_t^0| = k+2) \mathbb{E} [h(Z_t^0 + \eta + \delta_w) \mid |Z_t^0| = k+1]$ to the part in curly brackets.

Then, using 2.49) and (2.50), we find that

$$\begin{aligned}
 & \left| \mathbb{E} \left[h(Z_t^0 + \eta + \delta_z + \delta_w) - h(Z_t^0 + \eta + \delta_z) - h(Z_t^0 + \eta + \delta_w) + h(Z_t^0 + \eta) \right] \right| \\
 & \leq \sum_{k \geq -1} s_2(h) \left\{ P(|Z_t^0| = k) + P(|Z_t^0| = k + 2) \right\} / (k + 2 + |\eta|) \\
 & \quad + \left| \sum_{k \geq -1} \mathbb{E} \left[\frac{1}{2} h(Z_t^0 + \eta + \delta_z) + \frac{1}{2} h(Z_t^0 + \eta + \delta_w) \mid |Z_t^0| = k + 1 \right] \right. \\
 & \quad \left. \cdot \left\{ P(|Z_t^0| = k) - 2P(|Z_t^0| = k + 1) + P(|Z_t^0| = k + 2) \right\} \right| \\
 & \quad + P(|Z_t^0| = 0) |h(\eta)| \\
 & \leq s_2(h) \sum_{k \geq -1} \left\{ P(|Z_t^0| = k) + P(|Z_t^0| = k + 2) \right\} / (k + 2 + |\eta|) \\
 & \quad + \frac{s_2(h)}{2} \left\{ \sum_{k \geq -1} \left| P(|Z_t^0| = k) - 2P(|Z_t^0| = k + 1) + P(|Z_t^0| = k + 2) \right| \right. \\
 & \quad \left. + P(|Z_t^0| = 0) \right\}.
 \end{aligned}$$

Note that

$$\begin{aligned}
 & \sum_{k \geq -1} \frac{P(|Z_t^0| = k) + P(|Z_t^0| = k + 2)}{k + |\eta| + 2} \\
 & \leq \sum_{k \geq 0} \frac{P(|Z_t^0| = k)}{k + 2} + \sum_{k \geq 1} \frac{P(|Z_t^0| = k)}{k} \\
 & = \mathbb{E} \left[(|Z_t^0| + 2)^{-1} \right] + \mathbb{E} \left[(|Z_t^0|)^{-1} I_{\{|Z_t^0| \geq 1\}} \right] \\
 & \leq 3\mathbb{E} \left[(|Z_t^0| + 1)^{-1} \right]. \tag{2.56}
 \end{aligned}$$

Moreover, for $X \sim \text{Poi}(\nu)$,

$$P(X = k) - 2P(X = k - 1) + P(X = k - 2) = P(X = k) \left\{ (1 - \nu^{-1}k)^2 - \nu^{-2}k \right\},$$

for all $k \geq 0$. Since $|Z_t^0| \sim \text{Poi}(\lambda_t)$ with $\lambda_t := \lambda(1 - e^{-t})$ by Lemma 2.6.1, we

have

$$\begin{aligned}
 & \sum_{k \geq -1} |P(|Z_t^0| = k) - 2P(|Z_t^0| = k+1) + P(|Z_t^0| = k+2)| + P(|Z_t^0| = 0) \\
 &= \sum_{k \geq 0} |P(|Z_t^0| = k-2) - 2P(|Z_t^0| = k-1) + P(|Z_t^0| = k)| \\
 &= \sum_{k \geq 0} P(|Z_t^0| = k) \left| (1 - \lambda_t^{-1}|Z_t^0|)^2 - \lambda_t^{-2}|Z_t^0| \right| \\
 &\leq \mathbb{E} \left[(1 - \lambda_t^{-1}|Z_t^0|)^2 \right] + \mathbb{E} [\lambda_t^{-2}|Z_t^0|] = \frac{2}{\lambda_t}. \tag{2.57}
 \end{aligned}$$

It follows from (2.56), (2.57) and (2.52) that

$$\begin{aligned}
 & \left| \mathbb{E} [h(Z_t^0 + \eta + \delta_z + \delta_w) - h(Z_t^0 + \eta + \delta_z) - h(Z_t^0 + \eta + \delta_w) + h(Z_t^0 + \eta)] \right| \\
 &\leq s_2(h) \left\{ \frac{1}{\lambda_t} + \frac{3(1 - e^{-\lambda_t})}{\lambda_t} \right\} \leq \frac{4s_2(h)}{\lambda_t}.
 \end{aligned}$$

A more direct bound is given by $4 \sup_{\xi} |h(\xi)| = 2s_2(h)$. With these two estimates, we find, for any $\eta \in \mathcal{N}$, and for τ chosen such that $e^{-\tau} = 1 - 2\lambda^{-1}$,

$$\begin{aligned}
 & \int_0^\infty \frac{e^{-2t}}{s_2(h)} \left| \mathbb{E} [h(Z_t^0 + \eta + \delta_z + \delta_w) - h(Z_t^0 + \eta + \delta_z) - h(Z_t^0 + \eta + \delta_w) \right. \\
 & \quad \left. + h(Z_t^0 + \eta)] \right| dt \\
 &= \int_0^\tau 2e^{-2t} dt + \frac{4}{\lambda} \int_\tau^\infty \frac{e^{-2t}}{1 - e^{-t}} dt = 1 - e^{-2\tau} - \frac{4}{\lambda} e^{-\tau} - \frac{4}{\lambda} \log(1 - e^{-\tau}) \\
 &= 1 - \left(1 - \frac{2}{\lambda}\right)^2 - \frac{4}{\lambda} \left(1 - \frac{2}{\lambda}\right) - \frac{4}{\lambda} \log\left(\frac{2}{\lambda}\right) = \frac{4}{\lambda^2} + \frac{4}{\lambda} \log\left(\frac{\lambda}{2}\right).
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \Delta_2 \gamma &\leq s_2(h) \left\{ \frac{4}{\lambda^2} + \frac{4}{\lambda} \log\left(\frac{\lambda}{2}\right) \right\} \sum_{\eta \in \mathcal{N}} P(D_t = \eta) \leq s_2(h) \left\{ \frac{4}{\lambda^2} + \frac{4}{\lambda} \log\left(\frac{\lambda}{2}\right) \right\} \\
 &\leq \frac{2}{\lambda} \left\{ 1 + 2 \log_+ \left(\frac{\lambda}{2} \right) \right\} s_2(h), \quad \text{for all } \lambda \geq 2.
 \end{aligned}$$

□

Since the class of functions h is smaller than the class of functions considered for approximation in the total variation distance, the smoothness estimates are better: they have the desired property of decreasing with increasing λ . With the above lemmas, we are in a position to prove an analogue of Theorem 2.5.7 in the weaker d_2 -metric.

Theorem 2.6.5. *Suppose there exists a fixed measure ν on E and suppose that Ξ is a finite simple point process on E with finite mean measure λ and Janossy densities $\{j_m\}_{m \geq 0}$. Suppose the density μ of λ with respect to ν is given by (2.26). Let $\{N_z\}_{z \in E}$ be a neighbourhood structure satisfying (2.27). Then,*

$$\begin{aligned} d_2(\mathcal{L}(\Xi), \text{PRM}(\lambda)) \\ \leq \left\{ 1 \wedge \frac{2}{\lambda} \left(1 + 2 \log_+ \left(\frac{\lambda}{2} \right) \right) \right\} \left(\int_E \mathbb{E} \Xi(N_z) \mu(z) \nu(dz) \right. \\ \left. + \mathbb{E} \left[\int_E \Xi(N_z \setminus \{z\}) \Xi(dz) \right] \right) + \left\{ 1 \wedge \frac{1.65}{\sqrt{\lambda}} \right\} \int_E \mathbb{E} |g(z, \Xi^z) - \mu(z)| \nu(dz), \end{aligned}$$

where $\lambda = \lambda(E)$, and the conditional density $g(z, \Xi^z)$ at z given the configuration Ξ^z of Ξ outside N_z is defined in (2.28).

Proof. Let γ be defined as in Proposition 2.5.4 for a function $h \in \mathcal{H}$. By (2.41) and (2.43), it is sufficient to determine an upper bound on $|\mathbb{E}(\mathcal{A}\gamma)(\Xi)|/s_2(h)$, where \mathcal{A} is defined as in (2.30). From the proof of Theorem 2.5.7, we have that

$$\begin{aligned} |\mathbb{E}(\mathcal{A}\gamma)(\Xi)| &\leq \Delta_2 \gamma \left(\int_E \mathbb{E} \Xi(N_z) \mu(z) \nu(dz) + \mathbb{E} \left[\int_E \Xi(N_z \setminus \{z\}) \Xi(dz) \right] \right) \\ &\quad + \Delta_1 \gamma \int_E \mathbb{E} |g(z, \Xi^z) - \mu(z)| \nu(dz). \end{aligned}$$

The estimates from Lemmas 2.6.3 and 2.6.4 for $\Delta_1 \gamma$ and $\Delta_2 \gamma$, respectively, then immediately give the error bound for $d_2(\mathcal{L}(\Xi), \text{PRM}(\lambda))$. \square

In case we want to approximate the law of Ξ by a Poisson process with intensity measure $\tilde{\lambda} \neq \lambda$, we have to add an estimate for $d_2(\text{PRM}(\lambda), \text{PRM}(\tilde{\lambda}))$ to the error given by Theorem 2.6.5. To determine such an error estimate we first need Lemma 2.6.6.

Lemma 2.6.6. *Let λ and $\tilde{\lambda}$ be two finite measures over E such that $\lambda(E) = \tilde{\lambda}(E) = \lambda$. Let $\gamma : M_p(E) \rightarrow \mathbb{R}$ be defined as in Proposition 2.5.4, where*

$h : M_p(E) \rightarrow \mathbb{R}$ is any function in \mathcal{H} . Then, for any $\xi \in M_p(E)$,

$$\begin{aligned} \left| \int_E [\gamma(\xi + \delta_z) - \gamma(\xi)] (\lambda(dz) - \tilde{\lambda}(dz)) \right| \\ \leq s_2(h)(1 - e^{-\lambda}) \left(1 + \frac{\lambda}{|\xi| + 1} \right) d_1(\lambda, \tilde{\lambda}). \end{aligned}$$

Proof. For any $\xi \in M_p(E)$, define $\gamma_\xi : E \rightarrow \mathbb{R}$ by $\gamma_\xi(z) = \gamma(\xi + \delta_z) - \gamma(\xi)$. From the definition of d_1 in (2.37),

$$\begin{aligned} \left| \int_E [\gamma(\xi + \delta_z) - \gamma(\xi)] (\lambda(dz) - \tilde{\lambda}(dz)) \right| &= \left| \int_E \gamma_\xi d\lambda - \int_E \gamma_\xi d\tilde{\lambda} \right| \\ &\leq s_1(\gamma_\xi) \lambda d_1(\lambda, \tilde{\lambda}). \end{aligned} \quad (2.58)$$

In order to determine an upper bound on $s_1(\gamma_\xi)$, that is, on $|\gamma_\xi(z) - \gamma_\xi(w)|/|d_0(z, w)|$ for any choice of $z \neq w \in E$, let $Z = \{Z_t, t \in \mathbb{R}_+\}$ be an immigration-death process on E with initial point configuration ξ , i.e. realised under \mathbb{P}^ξ . Let T be an exponential random variable with rate 1 and independent of Z . The processes defined by $Z_t^z = Z_t + \delta_z I_{\{T > t\}}$ and $Z_t^w = Z_t + \delta_w I_{\{T > t\}}$ then have distributions $\mathbb{P}^{\xi + \delta_z}$ and $\mathbb{P}^{\xi + \delta_w}$, respectively, and, for any $z \neq w \in E$, $|\gamma_\xi(z) - \gamma_\xi(w)|$ equals

$$\begin{aligned} |\gamma(\xi + \delta_z) - \gamma(\xi + \delta_w)| &= \left| \int_0^\infty \mathbb{E}^\xi [\{h(Z_t^z) - h(Z_t^w)\} I_{\{T > t\}}] dt \right| \\ &\leq \int_0^\infty e^{-t} \mathbb{E}^\xi |h(Z_t + \delta_z) - h(Z_t + \delta_w)| dt \\ &\leq s_2(h) \int_0^\infty e^{-t} \mathbb{E}^\xi [d_1(Z_t + \delta_z, Z_t + \delta_w)] dt, \end{aligned} \quad (2.59)$$

where we used Lipschitz continuity of h in the last inequality. Now, note that $|Z_t + \delta_z| = |Z_t + \delta_w| = |Z_t| + 1$, and $|Z_t|$ is an immigration-death process on \mathbb{Z}_+ with initial number of points $|\xi|$. By (2.39) we thus have $\mathbb{E}^\xi [d_1(Z_t + \delta_z, Z_t + \delta_w)] = d_0(z, w) \mathbb{E}^\xi [(|Z_t| + 1)^{-1}]$ for any $t \in \mathbb{R}_+$. With Lemma 2.6.2, (2.59) then gives

$$|\gamma_\xi(z) - \gamma_\xi(w)| \leq s_2(h) d_0(z, w) (1 - e^{-\lambda}) \left(\frac{1}{\lambda} + \frac{1}{|\xi| + 1} \right),$$

for any $z \neq w \in E$, and therefore,

$$s_1(\gamma_\xi) \leq s_2(h)(1 - e^{-\lambda}) \left(\frac{1}{\lambda} + \frac{1}{|\xi| + 1} \right).$$

Use of this bound for $s_1(\gamma_\xi)$ in (2.58) completes the proof. \square

With the above lemma, we are in shape to determine an error estimate for the d_2 -distance between two Poisson processes with different mean measures λ and $\tilde{\lambda}$.

Proposition 2.6.7. *Let λ and $\tilde{\lambda}$ be two finite measures over E such that $\lambda(E) = \tilde{\lambda}(E) = \lambda$. Then*

$$d_2 \left(\text{PRM}(\lambda), \text{PRM}(\tilde{\lambda}) \right) \leq (1 - e^{-\lambda})(2 - e^{-\lambda})d_1(\lambda, \tilde{\lambda}).$$

Proof. Let $\Xi := \Xi_{\tilde{\lambda}} \sim \text{PRM}(\tilde{\lambda})$ and let $\Xi_{\lambda} \sim \text{PRM}(\lambda)$. Let \mathcal{A} be the generator of an immigration-death process with immigration intensity λ , unit per-capita death rate, and equilibrium distribution $\mathcal{L}(\Xi_{\lambda})$. By (2.43), $|\mathbb{E}h(\Xi) - \text{PRM}(\lambda)(h)| = |\text{PRM}(\tilde{\lambda})(h) - \text{PRM}(\lambda)(h)|$ equals $|\mathbb{E}(\mathcal{A}\gamma)(\Xi)|$. From the proof of Proposition 2.5.10 we know that

$$\mathbb{E}(\mathcal{A}\gamma)(\Xi) = \mathbb{E} \int_E [\gamma(\Xi + \delta_z) - \gamma(\Xi)] (\lambda(dz) - \tilde{\lambda}(dz)),$$

and thus

$$\begin{aligned} \frac{|\mathbb{E}(\mathcal{A}\gamma)(\Xi)|}{s_2(h)} &\leq \frac{1}{s_2(h)} \mathbb{E} \left| \int_E [\gamma(\Xi + \delta_z) - \gamma(\Xi)] (\lambda(dz) - \tilde{\lambda}(dz)) \right| \\ &\leq (1 - e^{-\lambda}) (1 + \lambda \mathbb{E} [(|\Xi| + 1)^{-1}]) d_1(\lambda, \tilde{\lambda}), \end{aligned}$$

where we used Lemma 2.6.6 for the second inequality. Finally, since $|\Xi| \sim \text{Poi}(\lambda)$, we have

$$\mathbb{E} [(|\Xi| + 1)^{-1}] = \sum_{k \geq 0} \frac{P(|\Xi| = k)}{k + 1} = e^{-\lambda} \sum_{k \geq 0} \frac{\lambda^k}{(k + 1)!} = \frac{e^{-\lambda}}{\lambda} \sum_{k \geq 1} \frac{\lambda^k}{k!} = \frac{1 - e^{-\lambda}}{\lambda},$$

and thus

$$1 + \lambda \mathbb{E} [(|\Xi| + 1)^{-1}] = 2 - e^{-\lambda}.$$

\square

3 Poisson and Poisson process approximation for univariate extremes

The tools that we established in Chapter 2 by way of the Stein-Chen method are now applied to problems from extreme value theory, where we restrict ourselves, for simplicity, to samples of i.i.d. univariate random variables. Section 3.1 relates extreme points to exceedances of thresholds. Since the number of extreme points follows a binomial distribution, the Stein-Chen method for Poisson approximation from Section 2.4 may be used in order to determine error estimates for the approximation by a Poisson distribution. We thereby establish bounds on the errors in the Kolmogorov distance involved in the approximation of the law of the maximum value by a so-called extreme value distribution. In particular, we delineate the different steps, as well as the respective error estimates arising from them, that are needed for the approximation. We present our results for the cases of random variables that follow exponential, Pareto, uniform, normal, Cauchy or geometric distributions. In Section 3.2 we generalise by introducing marked point processes of exceedances. Using results from Chapter 2.5 and 2.6, we determine and discuss bounds on the errors in the total variation distance (or the d_2 -distance if need be) for processes whose marks follow any of the distributions that we already treated in Section 3.1.

3.1 Poisson approximation for the number of extreme points and maxima of random variables

The first question is of course: what is an “extreme point”? It is an atypical value taken by a random variable. For a one-dimensional random variable X with state space $E \subseteq \mathbb{R}$ it is a value that exceeds a certain threshold, either towards the right or towards the left of the state space. Suppose we have

random variables X_1, \dots, X_n that are i.i.d. copies of X , and denote by F and \bar{F} the distribution and survival functions of X , respectively. In this section, we consider upper tail extremes, i.e. we call “extreme value” or “extreme point” (suggesting the language of point processes) a value in $(u_n, x_F] \cap E$ (or $[u_n, x_F] \cap E$), where $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$ is the right endpoint of F and u_n denotes a threshold that varies with the chosen sample size n (note that if $x_F = \infty$, then $(u_n, x_F] \cap E = (u_n, \infty)$). The number of extreme points is then given by

$$\sum_{i=1}^n I_{\{X_i > u_n\}} \sim \text{Bin}(n, \bar{F}(u_n)).$$

For a threshold u_n increasing with n , the probability $\bar{F}(u_n)$ of exceeding the threshold decreases towards 0. If $n\bar{F}(u_n) \rightarrow \lambda > 0$ as $n \rightarrow \infty$, the law of the number of extreme points converges to a Poisson distribution with mean λ . This implies that the number of points exceeding the threshold u_n is approximately distributed as $\text{Poi}(n\bar{F}(u_n))$. The Stein-Chen method for Poisson approximation from Section 2.4 provides us with the tools needed to investigate the sharpness of this approximation for each integer $n \geq 1$. Instead of counting points in $(u_n, x_F] \cap E$, we can count them in a more general set A that we suppose to be a measurable subset of E containing extreme values. Theorem 2.4.4 then gives the following result for the error in total variation that arises when approximating the law of $\sum_{i=1}^n I_{\{X_i \in A\}}$ by a Poisson distribution:

Theorem 3.1.1. *For each integer $n \geq 1$, let X, X_1, \dots, X_n be i.i.d. univariate random variables with state space $E \subseteq \mathbb{R}$. For a fixed set $A \in \mathcal{E} := \mathcal{B}(E)$, let $W_A := \sum_{i=1}^n I_{\{X_i \in A\}}$ denote the random number of points in A . Then,*

$$d_{TV}(\mathcal{L}(W_A), \text{Poi}(nP(X \in A))) \leq P(X \in A).$$

Proof. We apply Theorem 2.4.4 with $I_i := I_{\{X_i \in A\}}$ and $W := W_A$. Then $p_i \equiv P(X \in A)$, $\lambda = \mathbb{E}W_A = nP(X \in A)$, and the upper bound in (2.19) equals $P(X \in A)$. \square

For the case $A = A_n = (u_n, x_F] \cap E$, Theorem 3.1.1 amounts to

$$d_{TV}(\text{Bin}(n, \bar{F}(u_n)), \text{Poi}(n\bar{F}(u_n))) \leq \bar{F}(u_n). \quad (3.1)$$

This result is immediately applicable to all kinds of distributions F , and gives error bounds vanishing with $n \rightarrow \infty$ for suitably chosen thresholds u_n . It

can prominently be used to study the quality of asymptotic results given by classical extreme value theory, which establishes limit laws for maxima of i.i.d. random variables. The number of extreme points can be related to the maximum $X_{(n)} := \max_{1 \leq i \leq n} X_i$ of the random variables X_1, \dots, X_n by considering that

$$\left\{ \sum_{i=1}^n I_{\{X_i > u_n\}} = 0 \right\} = \{X_{(n)} \leq u_n\}.$$

Using the Poisson approximation to the binomial it is thus clearly possible to determine an approximation to the law of the maximum, and (3.1) in particular gives the error of this approximation:

$$|P(W_A = 0) - \text{Poi}(\text{EW}_A)\{0\}| = \left| P(X_{(n)} \leq u_n) - e^{-n\bar{F}(u_n)} \right| \leq \bar{F}(u_n). \quad (3.2)$$

Indeed, underlying classical extreme value theory is the following well-known limit result:

Theorem 3.1.2. (*Poisson approximation for maxima of i.i.d. rv's*) Let X_1, \dots, X_n be i.i.d. random variables with maximum $X_{(n)}$. For given $\tau \in [0, \infty]$ and a sequence $(u_n)_{n \geq 1}$ of real numbers, the following are equivalent:

$$n\bar{F}(u_n) \rightarrow \tau, \quad \text{as } n \rightarrow \infty, \quad (3.3)$$

$$P(X_{(n)} \leq u_n) \rightarrow e^{-\tau}, \quad \text{as } n \rightarrow \infty. \quad (3.4)$$

Proof. See, for instance, Proposition 3.1.1 in Embrechts et al. (1997). \square

More interesting than the approximation by $e^{-\tau}$ for a fixed value τ , or by $e^{-n\bar{F}(u_n)}$ which varies with the sample size n , would be the approximation by a non-degenerate distribution function that no longer depends on n . Such a distribution function may be found by subjecting the maximum to a *normalisation*, more precisely here, to a suitable affine transformation $u_n = a_n x + b_n$ for $x \in \mathbb{R}$, $a_n, b_n \in \mathbb{R}$ with $a_n > 0$. If we can indeed find non-degenerate limit distributions for maxima, then what are these? This question is answered by one of the most fundamental results of classical extreme value theory, the Fisher-Tippett theorem:

Theorem 3.1.3. (*Fisher-Tippett*) Let X_1, \dots, X_n be a sequence of i.i.d. random variables with maximum $X_{(n)}$. If there exist norming constants $a_n > 0$, $b_n \in \mathbb{R}$ and some non-degenerate distribution function H such that

$$\frac{X_{(n)} - b_n}{a_n} \xrightarrow{d} H, \quad (3.5)$$

then H belongs to the type of one of the three following distribution functions:

$$\begin{aligned} \text{Fréchet:} \quad \Phi_\alpha(x) &= \begin{cases} 0, & x \leq 0 \\ e^{-x^{-\alpha}}, & x > 0 \end{cases} & \alpha > 0. \\ \text{Weibull:} \quad \Psi_\alpha(x) &= \begin{cases} e^{-(-x)^\alpha}, & x < 0 \\ 1, & x \geq 0 \end{cases} & \alpha > 0. \\ \text{Gumbel:} \quad \Lambda(x) &= e^{-e^{-x}}, x \in \mathbb{R}. \end{aligned}$$

Proof. See, for instance, Proposition 0.3 in Resnick (1987). \square

The Fréchet, Weibull and Gumbel distributions are called *extreme value distributions*. If (3.5) holds, we say that F is in the *maximum domain of attraction* of H , which we denote by $F \in \text{MDA}(H)$. We summarise the normalisations and extremal limit results for a selection of well-known (continuous) distribution functions in the following proposition:

Proposition 3.1.4. *For any integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. random variables with cumulative distribution function F and maximum $X_{(n)}$.*

(a) (*Exponential distribution*) Let

$$F(y) = \begin{cases} 1 - e^{-\lambda y}, & y \geq 0, \\ 0 & y < 0, \end{cases}$$

with rate parameter $\lambda > 0$. Then, for all $x \in \mathbb{R}$,

$$P\left(X_{(n)} \leq \frac{x + \log n}{\lambda}\right) \longrightarrow e^{-e^{-x}} = \Lambda(x), \quad \text{as } n \rightarrow \infty.$$

(b) (*Pareto distribution*) Let

$$F(y) = \begin{cases} 1 - \left(\frac{\phi}{y}\right)^\alpha, & y \geq \phi, \\ 0, & y < \phi, \end{cases} \quad (3.6)$$

where $\alpha, \phi > 0$ denote the shape and scale parameters, respectively. Then, for all $x > 0$,

$$P\left(X_{(n)} \leq \phi n^{\frac{1}{\alpha}} x\right) \longrightarrow e^{-x^{-\alpha}} = \Phi_\alpha(x), \quad \text{as } n \rightarrow \infty.$$

(c) (Uniform distribution) Let

$$F(y) = \begin{cases} 0, & y < a, \\ \frac{y-a}{b-a}, & a \leq y < b, \\ 1, & y \geq b, \end{cases}$$

where $a, b \in \mathbb{R}$ with $a < b$. Then, for all $x < 0$,

$$P\left(X_{(n)} \leq \frac{(b-a)x}{n} + b\right) \longrightarrow e^x = \Psi_1(x), \quad \text{as } n \rightarrow \infty.$$

(d) (Standard normal distribution) Let $\Phi(y) := F(y) = \int_{-\infty}^y \varphi(t)dt$ and $\varphi(y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}}$ for $y \in \mathbb{R}$. Then, for all $x \in \mathbb{R}$,

$$P\left(X_{(n)} \leq \frac{x}{\sqrt{2\log n}} + \sqrt{2\log n} - \frac{\log \log n + \log 4\pi}{2\sqrt{2\log n}}\right) \longrightarrow e^{-e^{-x}} = \Lambda(x),$$

as $n \rightarrow \infty$.

(e) (Standard Cauchy distribution) Let

$$F(y) = \arctan(y)/\pi + 0.5 = \int_{-\infty}^y 1/\pi(1+t^2)dt$$

for $y \in \mathbb{R}$. Then, for any $x > 0$,

$$P\left(X_{(n)} \leq \frac{nx}{\pi}\right) \longrightarrow e^{-x^{-1}} = \Phi_1(x), \quad \text{as } n \rightarrow \infty.$$

Proof. For each case we use $y := y_n := a_n x + b_n$ and

$$\begin{aligned} P(X_{(n)} \leq a_n x + b_n) &= P(X_1 \leq a_n x + b_n, \dots, X_n \leq a_n x + b_n) \\ &= F^n(a_n x + b_n). \end{aligned} \tag{3.7}$$

(a) Fix any $x \in \mathbb{R}$. With $a_n = \lambda^{-1}$ and $b_n = \lambda^{-1} \log n$, we have, for any integer $n > e^{-x}$,

$$F^n(a_n x + b_n) = \left(1 - \frac{e^{-x}}{n}\right)^n \longrightarrow e^{-e^{-x}} = \Lambda(x), \quad \text{as } n \rightarrow \infty.$$

(b) Fix any $x > 0$. With $a_n = \phi n^{1/\alpha}$ and $b_n \equiv 0$, we have, for any integer $n > x^{-\alpha}$,

$$F^n(a_n x + b_n) = \left(1 - \frac{x^{-\alpha}}{n}\right)^n \rightarrow e^{-x^{-\alpha}} = \Phi_\alpha(x), \quad \text{as } n \rightarrow \infty.$$

(c) Fix any $x < 0$. With $a_n = (b - a)/n$ and $b_n \equiv b$, we have, for any integer $n > x$,

$$F^n(a_n x + b_n) = \left(1 + \frac{x}{n}\right)^n \rightarrow e^x = \Psi_1(x), \quad \text{as } n \rightarrow \infty.$$

(d) For any fixed $x \in \mathbb{R}$, we have $y = y_n = y_n(x) = a_n x + b_n \in \mathbb{R}$, where

$$a_n = \frac{1}{\sqrt{2 \log n}} \rightarrow 0 \quad \text{and} \quad b_n = \sqrt{2 \log n} - \frac{\log \log n + \log 4\pi}{2\sqrt{2 \log n}} \rightarrow \infty,$$

as $n \rightarrow \infty$. It follows that $y_n(x) \rightarrow \infty$ and $\bar{\Phi}(y_n) \rightarrow 0$ as $n \rightarrow \infty$. Hence,

$$\Phi^n(y_n) = (1 - \bar{\Phi}(y_n))^n = e^{n \log(1 - \bar{\Phi}(y_n))} \sim e^{-n \bar{\Phi}(y_n)} \sim e^{-\frac{n \varphi(y_n)}{y_n}},$$

as $y_n, n \rightarrow \infty$, where the first asymptotic equality is due to $\log(1 - z) \sim -z$ as $z \rightarrow 0$, and the second is the Mills ratio, i.e. $\bar{\Phi}(y_n) \sim y_n^{-1} \varphi(y_n)$ as $y_n \rightarrow \infty$. Furthermore, for x fixed,

$$\begin{aligned} & \frac{n \varphi(y_n)}{y_n} \\ &= e^{-x} \cdot \frac{\exp \left\{ -\frac{1}{8 \log n} [2x^2 - 2x(\log \log n + \log 4\pi) + (\log \log n + \log 4\pi)^2] \right\}}{1 + \frac{x - (\log \log n + \log 4\pi)/2}{2 \log n}} \end{aligned} \quad (3.8)$$

tends to e^{-x} as $n \rightarrow \infty$, and thus

$$\Phi^n(a_n x + b_n) \rightarrow e^{-e^{-x}} = \Lambda(x), \quad \text{as } n \rightarrow \infty.$$

(e) By l'Hôpital's rule,

$$\lim_{y \rightarrow \infty} \frac{\bar{F}(y)}{1/\pi y} = \lim_{y \rightarrow \infty} \frac{y^2}{1 + y^2} = 1.$$

With $a_n = n\pi^{-1}$ and $b_n \equiv 0$, we then find

$$F^n(a_n x + b_n) = \left(1 - \bar{F}\left(\frac{nx}{\pi}\right)\right)^n \sim \left(1 - \frac{x^{-1}}{n}\right)^n \rightarrow e^{-x^{-1}} = \Phi_1(x)$$

for all $x > 0$. □

The question we now ask is whether these limit results actually give good approximations for the laws of the maxima. Some results on convergence rates are given in Chapters 2.4 in Resnick (1987) and Leadbetter et al. (1983), respectively, and, for maxima of normals, in Hall (1979). We establish precise rates of convergence to extreme value distributions by using (3.2) along with suitable normalisations. In Section 3.1.1 we achieve this for each of the results from Proposition 3.1.4. Later, Section 3.1.2 discusses issues that may arise for distributions with discontinuities in the tail, and Section 3.1.3 offers a way to partially remedy these issues.

3.1.1 Maxima of continuous random variables

As demonstrated in Propositions 3.1.5, 3.1.6 and 3.1.7 below, it is relatively straightforward to establish rates of convergence for the maximum law of exponential, Pareto, and uniform random variables, using Theorem 3.1.1 and a suitable normalisation. For each of these cases the uniform error bound is of order $\log(n)/n \rightarrow 0$, as $n \rightarrow \infty$, thus providing a sharp approximation. The proofs for these three distributions are analogous, whereas the cases of standard normal and Cauchy random variables, treated in Propositions 3.1.8 and 3.1.9, respectively, are somewhat more involved. By choosing a non-linear normalisation for the uniform distribution, it is possible to expand the range of possible limiting distributions from the Weibull with parameter 1 to Weibull distributions with any parameter $\alpha > 0$. The choice $\alpha = 1$ restores the result from Proposition 3.1.4.

Proposition 3.1.5. (*Exponential distribution*) *For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. exponential random variables with parameter $\lambda > 0$. Then, for all $x \in \mathbb{R}$,*

$$\left| P\left(X_{(n)} \leq \frac{x + \log n}{\lambda}\right) - e^{-e^{-x}} \right| \leq \frac{\log n}{n} + \frac{1}{n} = O\left(\frac{\log n}{n}\right).$$

Proof. We apply Theorem 3.1.1 with $A = [y, \infty)$ for any choice of $y \geq 0$ to find

$$\left| P(X_{(n)} \leq y) - e^{-ne^{-\lambda y}} \right| \leq e^{-\lambda y}.$$

Plug in $y = \lambda^{-1}(x + \log n)$, possible for $x \geq -\log n$, since then $y \geq 0$, giving

$$\left| P\left(X_{(n)} \leq \frac{x + \log n}{\lambda}\right) - e^{-e^{-x}} \right| \leq \frac{e^{-x}}{n}.$$

In order to find a uniform bound for all $x \in \mathbb{R}$, choose $x_0 := x_{0n} := -\log \log n$. On the one hand, for all $x \geq x_0$, the error estimate $\exp(-x)/n$ is smaller than $\exp(-x_0)/n = \log(n)/n$. On the other hand, for all $x \leq x_0$,

$$P\left(X_{(n)} \leq \frac{x + \log n}{\lambda}\right) \leq P\left(X_{(n)} \leq \frac{x_0 + \log n}{\lambda}\right) \leq \frac{e^{-x_0}}{n} + e^{-e^{-x_0}},$$

since distribution functions are non-decreasing. This implies that

$$\left| P\left(X_{(n)} \leq \frac{x + \log n}{\lambda}\right) - e^{-e^{-x}} \right| \leq \frac{e^{-x_0}}{n} + e^{-e^{-x_0}} = \frac{\log n}{n} + \frac{1}{n}, \quad (3.9)$$

for all $x \leq x_0$. See Figure 3.1 for a sketch of the situation. We may use the upper bound in (3.9) for all $x \in \mathbb{R}$. \square

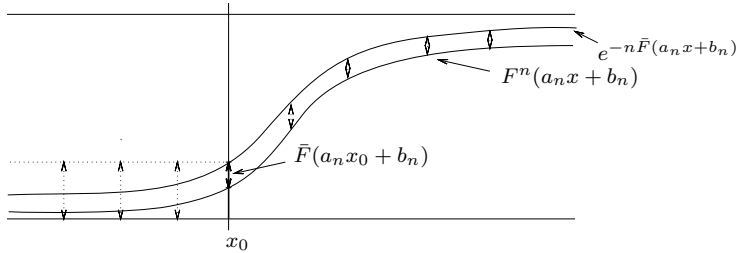


Figure 3.1: At all values $x \leq x_0$, the error cannot exceed the sum of the bound $\bar{F}(a_n x_0 + b_n)$ on the difference between the two distributions functions at x_0 and the height $\exp\{-n\bar{F}(a_n x_0 + b_n)\}$ of the approximating distribution.

Proposition 3.1.6. (*Pareto distribution*) For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. Pareto random variables with shape parameter $\alpha > 0$ and scale parameter $\phi > 0$. Then, for all $x > 0$,

$$\left| P\left(X_{(n)} \leq \phi n^{1/\alpha} x\right) - e^{-x^{-\alpha}} \right| \leq \frac{\log n}{n} + \frac{1}{n} = O\left(\frac{\log n}{n}\right).$$

Proof. We apply Theorem 3.1.1 with $A = [y, \infty)$ for any choice of $y \geq \phi$:

$$\left| P\left(X_{(n)} \leq y\right) - e^{-n\left(\frac{\phi}{y}\right)^\alpha} \right| \leq \left(\frac{\phi}{y}\right)^\alpha.$$

Plug in $y = \phi n^{1/\alpha} x$, possible for $x \geq n^{-1/\alpha}$, since then $y \geq \phi$, giving

$$\left| P\left(X_{(n)} \leq \phi n^{1/\alpha} x\right) - e^{-x^{-\alpha}} \right| \leq \frac{x^{-\alpha}}{n}.$$

In order to determine a uniform error bound, we first choose $x_0 := x_{0n} := (\log n)^{-1/\alpha}$. Then, for all $x \geq x_0$, we have $x^{-\alpha}/n \leq x_0^{-\alpha}/n = \log(n)/n$, whereas, for all $x \leq x_0$,

$$\left| P\left(X_{(n)} \leq \phi n^{1/\alpha} x\right) - e^{-x^{-\alpha}} \right| \leq \frac{x_0^{-\alpha}}{n} + e^{-x_0^{-\alpha}} = \frac{\log n}{n} + \frac{1}{n}.$$

This bound holds for all $x > 0$. □

Proposition 3.1.7. (*Uniform distribution*) For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. uniform random variables with parameters $a, b \in \mathbb{R}$, $a < b$, and let $\alpha > 0$. Then, for all $x < 0$,

$$\left| P\left(X_{(n)} \leq \frac{-(-x)^\alpha(b-a)}{n} + b\right) - e^{-(-x)^\alpha} \right| \leq \frac{\log n}{n} + \frac{1}{n} = O\left(\frac{\log n}{n}\right).$$

Proof. The proof is analogous to the proofs of Propositions 3.1.5 and 3.1.6, with the normalisation $y = -(-x)^\alpha(b-a)/n + b$. For all $y \in [a, b)$ or equivalently, for all $x \in [-n^{1/\alpha}, 0)$, we have

$$\left| P\left(X_{(n)} \leq \frac{-(-x)^\alpha(b-a)}{n} + b\right) - e^{-(-x)^\alpha} \right| \leq \frac{(-x)^\alpha}{n}.$$

With the choice $x_0 := x_{0n} := -(\log n)^{1/\alpha}$, we find the upper error bound

$$\frac{(-x_0)^\alpha}{n} + e^{-(-x_0)^\alpha} = \frac{\log n}{n} + \frac{1}{n},$$

which we may use for all $x < 0$. □

For the three examples above, the approximation by an extreme value distribution was effected in two steps: in the first step, we used Theorem 3.1.1 to approximate the maximum law by a Poisson probability mass function at 0, and in the second step, a suitable normalisation to transform this Poisson into the required extreme value distribution. For each of these examples the total error estimate is of order $\log(n)/n$, thus providing a sharp approximation as n increases. Moreover, the total error estimate is of the same order as the error estimate for the first step, which means that the principal part of the error in the approximation by an extreme value distribution arises from basic Poisson approximation, whereas the normalisation is negligible.

For the maximum of i.i.d. standard normals, the situation is somewhat different. As remarked already by Fisher and Tippett (1928), and later by Hall (1979) and Leadbetter et al. (1983), convergence of the maximum law of normals to the Gumbel distribution is extremely slow. Hall (1979) showed that with the choice of norming constants (3.12), the rate of convergence is not better than $(\log \log n)^2 / \log n$. Moreover, he showed that if the norming constants a_n and b_n were chosen as solutions to

$$\frac{n\varphi(b_n)}{b_n} = 1, \quad a_n = b_n^{-1},$$

then

$$\frac{C_1}{\log n} \leq \sup_{x \in \mathbb{R}} \left| P(X_{(n)} \leq a_n x + b_n) - e^{-e^{-x}} \right| \leq \frac{C_2}{\log n}$$

for constants $C_1, C_2 > 0$, and that the rate of convergence cannot be improved by choosing different norming constants.

Proposition 3.1.8 below delineates the different steps needed for the approximation by the Gumbel distribution and gives error estimates for each step. Though the first step, which is basic Poisson approximation, gives an error that is only of order $\log(n)/n$, the subsequent steps needed for the approximation by a Gumbel distribution give bigger error estimates. The next step after Poisson approximation uses the *Mills ratio*:

$$\bar{\Phi}(y) \sim \frac{\varphi(y)}{y}, \quad \text{as } y \rightarrow \infty.$$

As shown below we estimate the error arising from this step, i.e. the error of the approximation of $\exp\{-n\bar{\Phi}(y)\}$ by $\exp\{-n\varphi(y)/y\}$, to be of order $1/\log n$, and thereby substantially bigger than the one from Poisson approximation. In a last

step, the normalisation needed to transform $\exp\{-n\varphi(y)/y\}$ into the required Gumbel distribution gives rise to an even bigger error, which we estimate to be of size $(\log \log n)^2 / \log n$.

Proposition 3.1.8. *(Normal distribution) For each integer $n \geq 2$, let X_1, \dots, X_n be i.i.d. standard normal random variables with distribution function Φ and probability density function φ . Then,*

(a) *(Basic Poisson approximation) For all $y \in \mathbb{R}$,*

$$\left| P(X_{(n)} \leq y) - e^{-n\bar{\Phi}(y)} \right| \leq \frac{\log n}{n} + \frac{1}{n} =: \delta_{\text{PoiAppr}} = O\left(\frac{\log n}{n}\right). \quad (3.10)$$

(b) *(Approximation using the Mills ratio) For each integer $n \geq 21$ and for all $y \in \mathbb{R}$,*

$$\begin{aligned} \left| P(X_{(n)} \leq y) - e^{-\frac{n\varphi(y)}{y}} \right| \\ \leq \delta_{\text{PoiAppr}} + \frac{1}{2\log n} + e^{-0.1\sqrt{\log n}} =: \delta_{\text{Mills}} = O\left(\frac{1}{\log n}\right). \end{aligned} \quad (3.11)$$

(c) *(Approximation by a standard Gumbel distribution) For each integer $n \geq 21$ and for all $x \in \mathbb{R}$,*

$$\begin{aligned} \left| P(X_{(n)} \leq a_n x + b_n) - e^{-e^{-x}} \right| &\leq \delta_{\text{Mills}} + \frac{69(\log \log n + \log 4\pi)^2}{\log n} \\ &= O\left(\frac{\log^2 \log n}{\log n}\right), \end{aligned}$$

where

$$a_n = \frac{1}{\sqrt{2\log n}}, \quad b_n = \sqrt{2\log n} - \frac{\log \log n + \log 4\pi}{2\sqrt{2\log n}}. \quad (3.12)$$

Proof. (a) We apply Theorem 3.1.1 with $A = [y, \infty)$ for any choice of $y \in \mathbb{R}$:

$$\left| P(X_{(n)} \leq y) - e^{-n\bar{\Phi}(y)} \right| \leq \bar{\Phi}(y). \quad (3.13)$$

For a uniform error bound, choose $y_0 := y_{0n} := \Phi^{-1}(1 - \log(n)/n)$. Then, for all $y \geq y_0$, we have $\bar{\Phi}(y) \leq \bar{\Phi}(y_0) = \log(n)/n$, whereas for $y \leq y_0$,

$$\left| P(X_{(n)} \leq y) - e^{-n\bar{\Phi}(y)} \right| \leq \bar{\Phi}(y_0) + e^{-n\bar{\Phi}(y_0)} = \frac{\log n}{n} + \frac{1}{n},$$

and we may use this bound for all $y \in \mathbb{R}$.

(b) By adding and subtracting $\exp\{-n\varphi(y)/y\}$ into (3.10), we find, for all $y \in \mathbb{R}$,

$$\left| P(X_{(n)} \leq y) - e^{-\frac{n\varphi(y)}{y}} \right| \leq \frac{\log n}{n} + \frac{1}{n} + \left| e^{-n\bar{\Phi}(y)} - e^{-\frac{n\varphi(y)}{y}} \right|,$$

and we need to determine a uniform bound for $\left| \exp\{-n\bar{\Phi}(y)\} - \exp\left\{-\frac{n\varphi(y)}{y}\right\} \right|$. Suppose first that $y > 0$. With the two consecutive changes of variables $z := t - y$ and $w := yz$, we obtain

$$\begin{aligned} \bar{\Phi}(y) &= \frac{1}{\sqrt{2\pi}} \int_y^\infty e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(z+y)^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \int_0^\infty e^{-z^2/2} \cdot e^{-zy} dz = \frac{\varphi(y)}{y} \int_0^\infty e^{-w^2/2y^2} \cdot e^{-w} dw. \end{aligned}$$

With $1 - x \leq e^{-x} \leq 1$ for $x = w^2/2y^2$, $\int_0^\infty e^{-w} dw = 1$ and $\int_0^\infty w^2 e^{-w} dw = 2$, we get

$$1 - \frac{1}{y^2} \leq \int_0^\infty e^{-w^2/2y^2} \cdot e^{-w} dw \leq 1,$$

and thereby

$$\frac{\varphi(y)}{y} - \frac{\varphi(y)}{y^3} \leq \bar{\Phi}(y) \leq \frac{\varphi(y)}{y}, \quad \text{for all } y > 0. \quad (3.14)$$

It then follows that for all $y \geq 2$,

$$\begin{aligned} \left| e^{-n\bar{\Phi}(y)} - e^{-\frac{n\varphi(y)}{y}} \right| &= e^{-n\bar{\Phi}(y)} - e^{-\frac{n\varphi(y)}{y}} = e^{-n\bar{\Phi}(y)} \left\{ 1 - e^{-n\left(\frac{\varphi(y)}{y} - \bar{\Phi}(y)\right)} \right\} \\ &\leq ne^{-n\bar{\Phi}(y)} \left(\frac{\varphi(y)}{y} - \bar{\Phi}(y) \right) \leq \frac{n\varphi(y)}{y^3} e^{-\frac{n\varphi(y)}{y} \left(1 - \frac{1}{y^2}\right)} \\ &\leq \frac{n\varphi(y)}{y^3} e^{-\frac{3n\varphi(y)}{4y}} \leq \frac{4}{3ey^2}, \end{aligned} \quad (3.15)$$

where we used $1 - e^{-z} \leq z$, (3.14) and $y \geq 2$ in the first, second and third inequalities, respectively, as well as $z \exp\{-(3/4)z\} \leq 4/(3e)$ for all $z \in \mathbb{R}$, for the last inequality. In order to determine a uniform bound on the error between $\exp\{-n\bar{\Phi}(y)\}$ and $\exp\{-n\varphi(y)/y\}$, choose $y_1 := y_{1n} := \sqrt{2 \log n} -$

$\log \log(n)/\sqrt{2 \log n}$. Note that $y_{1n} \geq 2$ for all $n \geq 21$. Since the error bound in (3.15) decreases for increasing y , we have

$$\begin{aligned} \frac{4}{3ey^2} &\leq \frac{4}{3ey_1^2} = \frac{2}{3e \log n} \left[1 - \frac{\log \log n}{\log n} + \frac{\log^2 \log n}{4 \log^2 n} \right]^{-1} \\ &\leq \frac{2}{3e \log n} \left[1 - \frac{\log \log n}{\log n} \right]^{-1} \leq \frac{4}{3e \log n} \leq \frac{1}{2 \log n} \end{aligned} \quad (3.16)$$

for all $y \geq y_1$, where we used that $\log \log(n)/\log n \leq 1/2$ for all $n \geq 1$. On the other hand, for all $y \leq y_1$,

$$\left| e^{-n\bar{\Phi}(y)} - e^{-\frac{n\varphi(y)}{y}} \right| \leq \frac{4}{3ey_1^2} + e^{-\frac{n\varphi(y_1)}{y_1}}. \quad (3.17)$$

This gives an upper bound that we may use for all $y \in \mathbb{R}$. We have

$$\begin{aligned} \varphi(y_1) &= \frac{e^{-y_1^2/2}}{\sqrt{2\pi}} = \frac{\log n}{n} \cdot \frac{e^{-\frac{\log^2 \log n}{4 \log n}}}{\sqrt{2\pi}}, \\ e^{-\frac{n\varphi(y_1)}{y_1}} &= \exp \left\{ -\sqrt{\log n} \cdot \frac{e^{-\frac{\log^2 \log n}{4 \log n}}}{2\sqrt{\pi} \left(1 - \frac{\log \log n}{2 \log n} \right)} \right\} \leq e^{-0.1\sqrt{\log n}}, \end{aligned} \quad (3.18)$$

where we used $(1 - (\log \log n)/(2 \log n))^{-1} \geq 1$ and $(2\sqrt{\pi})^{-1} \exp\{-(\log^2 \log n)/(4 \log n)\} \geq (2e\sqrt{\pi})^{-1} \geq 0.1$. It follows from (3.16)-(3.18) that for each integer $n \geq 21$,

$$\left| e^{-n\bar{\Phi}(y)} - e^{-\frac{n\varphi(y)}{y}} \right| \leq \frac{1}{2 \log n} + e^{-0.1\sqrt{\log n}}, \quad \text{for all } y \in \mathbb{R}. \quad (3.19)$$

(c) With the normalisation

$$y := y(x) = a_n x + b_n = \frac{x}{\sqrt{2 \log n}} + \sqrt{2 \log n} - \frac{\log \log n + \log 4\pi}{2\sqrt{2 \log n}}$$

we obtain $\frac{n\varphi(y)}{y} = e^{-x} f_n(x)$, where

$$f_n(x) = \frac{\exp \left\{ -\frac{1}{8 \log n} [2x^2 - 2x(\log \log n + \log 4\pi) + (\log \log n + \log 4\pi)^2] \right\}}{1 + \frac{x(\log \log n + \log 4\pi)/2}{2 \log n}}. \quad (3.20)$$

By adding and subtracting $\exp\{-e^{-x}\}$ into (3.11), we find

$$\begin{aligned}
 \left| P(X_{(n)} \leq a_n x + b_n) - e^{-e^{-x}} \right| &\leq \delta_{\text{Mills}} + \left| e^{-\frac{n\varphi(y)}{y}} - e^{-e^{-x}} \right| \\
 &\leq \delta_{\text{Mills}} + e^{-\min\{\frac{n\varphi(y)}{y}, e^{-x}\}} \cdot \left| \frac{n\varphi(y)}{y} - e^{-x} \right| \\
 &\leq \delta_{\text{Mills}} + e^{-\min\{e^{-x}f_n(x), e^{-x}\}} e^{-x} |f_n(x) - 1|,
 \end{aligned} \tag{3.21}$$

and we need to determine a uniform bound on the new error term in (3.21), for all $x \in \mathbb{R}$. Note that the function $2x^2 - 2ax + a^2$ takes the minimal value $a^2/2$ at $x = a/2$. Thus

$$2x^2 - 2x(\log \log n + \log 4\pi) + (\log \log n + \log 4\pi)^2 \geq \frac{1}{2}(\log \log n + \log 4\pi)^2,$$

which is strictly positive for $n \geq 2$, implying that

$$f_n(x) \leq \left[1 + \frac{x - (\log \log n + \log 4\pi)/2}{2 \log n} \right]^{-1}. \tag{3.22}$$

Suppose first that $x \geq \log \log n$. Then

$$e^{-\min\{e^{-x}f_n(x), e^{-x}\}} \leq 1, \quad e^{-x} \leq \frac{1}{\log n},$$

and

$$f_n(x) \leq \left[1 + \frac{\log \log n + \log 4\pi}{4 \log n} \right]^{-1} \leq 1, \quad |f_n(x) - 1| \leq f_n(x) + 1 \leq 2.$$

Thus,

$$\left| P(X_{(n)} \leq a_n x + b_n) - e^{-e^{-x}} \right| \leq \delta_{\text{Mills}} + \frac{2}{\log n}, \quad \text{for all } x \geq \log \log n. \tag{3.23}$$

Now suppose that $-\log \log n \leq x \leq \log \log n$. Then, on the one hand

$$\begin{aligned}
 f_n(x) &\leq \left[1 - \frac{3 \log \log n + \log 4\pi}{4 \log n} \right]^{-1} \\
 &\leq 1 + \frac{3 \log \log n + \log 4\pi}{4 \log n} \cdot \left[1 - \frac{3 \log \log n + \log 4\pi}{4 \log n} \right]^{-2} \\
 &\leq 1 + \frac{6 \log \log n + 2 \log 4\pi}{\log n},
 \end{aligned} \tag{3.24}$$

where the second inequality uses Taylor expansion about 0, and the third bounds the squared term by the constant 8, for all $n \geq 1$. Moreover, $e^{-x} \cdot \exp\{-e^{-x}\}$ has the global maximum e^{-1} . Thus, if $f_n(x) \geq 1$, the error in (3.21) is

$$\begin{aligned} \delta_{\text{Mills}} + e^{-e^{-x}} \cdot e^{-x} \cdot (f_n(x) - 1) &\leq \delta_{\text{Mills}} + e^{-1} \cdot \frac{6 \log \log n + 2 \log 4\pi}{\log n} \\ &\leq \delta_{\text{Mills}} + \frac{3 \log \log n + \log 4\pi}{\log n}. \end{aligned} \quad (3.25)$$

On the other hand, first note that for $-\log \log n \leq x \leq \log \log n$, and for all $n \geq 3$,

$$\begin{aligned} \left[1 + \frac{x - (\log \log n + \log 4\pi)/2}{2 \log n}\right]^{-1} &\geq \left[1 + \frac{\log \log n - \log 4\pi}{4 \log n}\right]^{-1} \\ &\geq 1 - \frac{\log \log n - \log 4\pi}{4 \log n}, \end{aligned}$$

where

$$0 \leq \frac{\log \log n - \log 4\pi}{4 \log n} \leq 0.9, \quad \text{for all } n \geq 2.$$

Moreover, note that

$$\begin{aligned} &2x^2 - 2x(\log \log n + \log 4\pi) + (\log \log n + \log 4\pi)^2 \\ &\leq 2 \log^2 \log n + 2 \log \log n (\log \log n + \log 4\pi) + (\log \log n + \log 4\pi)^2 \\ &\leq 5(\log \log n + \log 4\pi)^2, \end{aligned}$$

and that $5(\log \log n + \log 4\pi)^2 / (8 \log n) \leq 3$ for $n \geq 12$. Therefore,

$$\begin{aligned} f_n(x) &\geq \exp \left\{ -\frac{5(\log \log n + \log 4\pi)^2}{8 \log n} \right\} \left(1 - \frac{\log \log n - \log 4\pi}{4 \log n} \right) \\ &= \exp \left\{ -\frac{5(\log \log n + \log 4\pi)^2}{8 \log n} + \log \left(1 - \frac{\log \log n - \log 4\pi}{4 \log n} \right) \right\} \end{aligned} \quad (3.26)$$

$$\geq 0.1e^{-3} \geq 0.004. \quad (3.27)$$

Using (3.26) and $1 - e^{-z} \leq z$ for all $z \geq 0$, we obtain

$$\begin{aligned} 1 - f_n(x) &\leq \frac{5(\log \log n + \log 4\pi)^2}{8 \log n} - \log \left(1 - \frac{\log \log n - \log 4\pi}{4 \log n} \right) \\ &\leq \frac{5(\log \log n + \log 4\pi)^2}{8 \log n} + \frac{5(\log \log n - \log 4\pi)}{2 \log n} \\ &\leq \frac{3(\log \log n + \log 4\pi)^2}{4 \log n}, \end{aligned}$$

where we used $-\log(1-z) \leq 10z$ for $0 \leq z \leq 0.9$ for the second inequality, and $\log \log n - \log 4\pi \leq (\log \log n + \log 4\pi)^2/20$ for the third inequality. Furthermore, due to (3.27) $e^{-x} \cdot \exp\{-f_n(x)e^{-x}\} \leq e^{-x} \cdot \exp\{-0.004e^{-x}\} \leq 250/e$, for all $x \in \mathbb{R}$. Thus, if $f_n(x) \leq 1$, the error in (3.21) is

$$\begin{aligned} &\delta_{\text{Mills}} + e^{-e^{-x}f_n(x)} \cdot e^{-x}(1 - f_n(x)) \\ &\leq \delta_{\text{Mills}} + \frac{750}{4e} \cdot \frac{(\log \log n + \log 4\pi)^2}{\log n} \\ &\leq \delta_{\text{Mills}} + \frac{69(\log \log n + \log 4\pi)^2}{\log n}, \quad \text{for all } n \geq 12. \end{aligned} \quad (3.28)$$

All in all, (3.25) and (3.28) give, for all $n \geq 12$ and $-\log \log n \leq x \leq \log \log n$,

$$\left| P(X_{(n)} \leq a_n x + b_n) - e^{-e^{-x}} \right| \leq \delta_{\text{Mills}} + \frac{69(\log \log n + \log 4\pi)^2}{\log n}. \quad (3.29)$$

Lastly, suppose that $x \leq -\log \log n$. Since $a_n x + b_n \leq -a_n \log \log n + b_n$ and since $\exp\{-n\varphi(y)/y\}$ and $\exp\{-e^{-x}\}$ are non-decreasing, we have

$$\left| e^{-\frac{n\varphi(y)}{y}} - e^{-e^{-x}} \right| \leq e^{-\frac{n\varphi(y)}{y}} + e^{-e^{-x}} \leq e^{-\frac{n\varphi(-a_n \log \log n + b_n)}{-a_n \log \log n + b_n}} + \frac{1}{n},$$

where $\frac{n\varphi(-a_n \log \log n + b_n)}{-a_n \log \log n + b_n} = f_n(-\log \log n) \log n$, and

$$f_n(-\log \log n) \geq \frac{\exp\left\{-\frac{5(\log \log n + \log 4\pi)^2}{8 \log n}\right\}}{1 - \frac{3 \log \log n + \log 4\pi}{4 \log n}} \geq \exp\left\{-\frac{5(\log \log n + \log 4\pi)^2}{8 \log n}\right\},$$

which is bigger than $e^{-3} \geq 0.04$ for all $n \geq 12$. Thus,

$$e^{-\frac{n\varphi(-a_n \log \log n + b_n)}{-a_n \log \log n + b_n}} \leq e^{-0.04 \log n},$$

and, for all $n \geq 12$ and $x \leq -\log \log n$, we have

$$\left| P(X_{(n)} \leq a_n x + b_n) - e^{-e^{-x}} \right| \leq \delta_{\text{Mills}} + e^{-0.04 \log n} + \frac{1}{n}. \quad (3.30)$$

For a bound for all $x \in \mathbb{R}$ and for all $n \geq 12$, we choose the maximum of the bounds in (3.23), (3.29) and (3.30):

$$\left| P(X_{(n)} \leq a_n x + b_n) - e^{-e^{-x}} \right| \leq \delta_{\text{Mills}} + \frac{69(\log \log n + \log 4\pi)^2}{\log n}.$$

□

In order to approximate the law of the maximum of standard Cauchy random variables by a Fréchet distribution, we need an intermediate step (comparable to the Mills ratio for the case of maxima of normals). More precisely, we use that

$$\overline{F}(y) \sim \frac{1}{\pi y}, \quad \text{as } y \rightarrow \infty,$$

where $\overline{F}(y)$ denotes the survival function of the Cauchy distribution. Contrary to the case of maxima of normals, the normalisation here does not produce any additional error and the total error is of the same order as the error that we obtain in the first step from basic Poisson approximation, that is, of order $\log(n)/n$.

Proposition 3.1.9. *(Standard Cauchy distribution) For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. standard Cauchy random variables with distribution function $F(y) = \arctan(y)/\pi + 0.5$ and density $f(y) = 1/\pi(1+y^2)$, for $y \in \mathbb{R}$. Then:*

(a) *(Basic Poisson approximation) For all $y > 0$,*

$$\left| P(X_{(n)} \leq y) - e^{-n\overline{F}(y)} \right| \leq \frac{\log n}{n} + \frac{1.74}{n} = O\left(\frac{\log n}{n}\right).$$

(b) *(Approximation by a standard Fréchet distribution) For all $x > 0$,*

$$\left| P\left(X_{(n)} \leq \frac{nx}{\pi}\right) - e^{-x^{-1}} \right| \leq \frac{\log n}{n} + \frac{\pi^2 \log^3 n}{3n^2} + \frac{1}{n} = O\left(\frac{\log n}{n}\right). \quad (3.31)$$

Proof. We first determine an upper and lower bound on

$$\bar{F}(y) = 0.5 - \arctan(y)/\pi$$

by setting $z = 1/y$ and performing a Taylor expansion of $g(z) := \bar{F}(1/z)$ about $z = 0$:

$$g(z) = \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{1}{z}\right) = \frac{z}{\pi} - \frac{z^3}{3\pi} + \frac{z^5}{5\pi} - \frac{z^7}{7\pi} + \dots,$$

Here,

$$\begin{aligned} g(0+) &= \lim_{z \rightarrow 0+} f(z) = 0, \quad g'(z) = \frac{1}{\pi(1+z^2)}, \\ g''(z) &= -\frac{2z}{\pi(1+z^2)^2}, \quad g'''(z) = \frac{6z^2-2}{\pi(1+z^2)^3}. \end{aligned}$$

Since

$$\frac{z}{\pi} + \frac{z^3}{3!} \min_{0 \leq \xi \leq z} g'''(\xi) \leq g(z) \leq z \max_{0 \leq \xi \leq z} |g'(\xi)|,$$

where $\max_{0 \leq \xi \leq z} |g'(\xi)| = 1/\pi$ and $\min_{0 \leq \xi \leq z} g'''(\xi) = g'''(0) = -2/\pi$, we find

$$\frac{z}{\pi} - \frac{z^3}{3\pi} \leq g(z) \leq \frac{z}{\pi},$$

and thereby

$$\frac{1}{\pi y} - \frac{1}{3\pi y^3} \leq \bar{F}(y) \leq \frac{1}{\pi y}, \quad (3.32)$$

which holds for all $y > 0$.

(a) We apply Theorem 3.1.1 with $A = [y, \infty)$ for any choice of $y > 0$:

$$\left| P(X_{(n)} \leq y) - e^{-n\bar{F}(y)} \right| \leq \bar{F}(y) \leq \frac{1}{\pi y}, \quad (3.33)$$

where we used (3.32) for the second inequality. In order to determine a uniform error bound, first choose $y_0 := y_{0n} := n/\pi \log n$. Then, for all $y \geq y_0$, we have $1/\pi y \leq 1/\pi y_0 \leq \log(n)/n$, whereas, for $0 < y \leq y_0$, we may bound the error by further adding the approximating distribution function at y_0 , i.e. $\exp\{-n\bar{F}(y_0)\}$, to $1/\pi y_0$. Therefore,

$$\left| P(X_{(n)} \leq y) - e^{-n\bar{F}(y)} \right| \leq \frac{1}{\pi y_0} + e^{-n\bar{F}(y_0)} \leq \frac{\log n}{n} + \frac{1}{n} \exp \left\{ \frac{\pi^2 \log^3 n}{3n^2} \right\},$$

for all $y > 0$, where we used (3.32) to estimate $\exp\{-n\bar{F}(y_0)\}$. By noting that the function $\exp\{\pi^2 \log^3 z / 3z^2\}$ takes on its maximum $\exp\{9\pi^2/8e^3\} \approx 1.7381$ at $z = e^{3/2}$, we obtain the uniform bound $\log(n)/n + 1.74/n$.

(b) By adding and subtracting $\exp\{-n/\pi y\}$ into (3.33) and noting that $\exp\{-n\bar{F}(y)\} - \exp\{-n/\pi y\} \geq 0$, we obtain

$$\left| P(X_{(n)} \leq y) - e^{-\frac{n}{\pi y}} \right| \leq \frac{1}{\pi y} + e^{-n\bar{F}(y)} - e^{-\frac{n}{\pi y}},$$

where, using (3.32) for the last inequality,

$$\begin{aligned} e^{-n\bar{F}(y)} - e^{-\frac{n}{\pi y}} &= e^{-n\bar{F}(y)} \left\{ 1 - e^{-n(\frac{1}{\pi y} - \bar{F}(y))} \right\} \leq n e^{-n\bar{F}(y)} \left(\frac{1}{\pi y} - \bar{F}(y) \right) \\ &\leq n \left(\frac{1}{\pi y} - \bar{F}(y) \right) \leq \frac{n}{3\pi y^3}. \end{aligned}$$

Now plug in $y = nx/\pi$. Since $y > 0$, we have $x > 0$, and obtain

$$\begin{aligned} \left| P(X_{(n)} \leq y) - e^{-\frac{n}{\pi y}} \right| &= \left| P\left(X_{(n)} \leq \frac{nx}{\pi}\right) - e^{-x^{-1}} \right| \\ &\leq \frac{1}{\pi y} + \frac{n}{3\pi y^3} = \frac{1}{nx} + \frac{\pi^2}{3n^2 x^3}. \end{aligned}$$

To find a uniform error bound, choose $x_0 := x_{0n} := 1/\log n$ (which is equivalent to the choice of y_0 we had in (a)). For all $x \geq x_0$, we have

$$\frac{1}{nx} + \frac{\pi^2}{3n^2 x^3} \leq \frac{1}{nx_0} + \frac{\pi^2}{3n^2 x_0^3} = \frac{\log n}{n} + \frac{\pi^2 \log^3 n}{3n^2},$$

whereas for all $0 < x \leq x_0$,

$$\left| P\left(X_{(n)} \leq \frac{nx}{\pi}\right) - e^{-x^{-1}} \right| \leq \frac{1}{nx_0} + \frac{\pi^2}{3n^2 x_0^3} + e^{-x_0^{-1}} = \frac{\log n}{n} + \frac{\pi^2 \log^3 n}{3n^2} + \frac{1}{n}.$$

The latter bound clearly holds for all $x > 0$. □

3.1.2 The problem for maxima of discrete random variables

So far, we have only studied maxima of random variables with continuous distribution function F . As we will see below, it turns out that for some

well-known discrete distributions functions, no non-degenerate limit law H as mentioned in Theorem 3.1.3 may be found. So what are the conditions for the existence of non-degenerate limit laws? The following corollaries of Theorem 3.1.2 address this question. We again use the notation x_F for the right endpoint of the distribution function F of the i.i.d. random variables X_1, \dots, X_n , and note that $F(x) < 1$ for all $x < x_F$ and $F(x_F) = 1$ for all $x \geq x_F$. Moreover, we use the notation $F(x-) := \lim_{h \downarrow 0} F(x-h)$, for all $x \in \mathbb{R}$.

Corollary 3.1.10. (i) $X_{(n)} \rightarrow x_F$ with probability one as $n \rightarrow \infty$.
(ii) Suppose that $x_F < \infty$ and $F(x_F-) < 1$. Then, for every sequence $(u_n)_{n \geq 1}$ such that $P[X_{(n)} \leq u_n] \rightarrow \rho$, as $n \rightarrow \infty$, we either have $\rho = 0$ or $\rho = 1$.

Proof. See, for example, Corollary 1.5.2 in Leadbetter et al. (1983). \square

Thus, for X_1, \dots, X_n having common distribution function F with finite right endpoint $x_F < \infty$ and a jump at x_F , i.e. $F(x_F-) < 1 = F(x_F)$, it follows that if $P(X_{(n)} \leq u_n) = P(X_{(n)} \leq a_n x + b_n) \rightarrow \rho = H(x)$ for a sequence $u_n = a_n x + b_n$, then $H(x) = 0$ or 1 for each x , so that H is degenerate.

Example 3.1.11. (Binomial distribution) Let $X \sim \text{Bin}(m, p)$ for $m \in \mathbb{N}$ and $p \in (0, 1)$. Since $P(X = m) = 1$, the right endpoint x_F is given by $m < \infty$. Moreover, $F(x_F-) = F(m-1) = 1 - p^m < 1$. Thus, by Corollary 3.1.10, there exists no non-degenerate limit distribution for the maximum of binomials.

More commonly, the existence of non-degenerate limit laws is impossible due to the following corollary of Theorem 3.1.2, which is valid for distribution functions F with right endpoint $x_F \leq \infty$.

Theorem 3.1.12. Let X_1, \dots, X_n be i.i.d. random variables with common distribution function F and let $\tau \in (0, \infty)$. Then there exists a sequence $(u_n)_{n \geq 1}$ satisfying $n\bar{F}(u_n) \rightarrow \tau$ as $n \rightarrow \infty$ if and only if

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x)}{\bar{F}(x-)} = 1, \quad (3.34)$$

or equivalently, if and only if

$$\lim_{x \uparrow x_F} \frac{p(x)}{\bar{F}(x-)} = 0, \quad (3.35)$$

where $p(x) = F(x) - F(x-)$ and $p(x)/\bar{F}(x-)$ denotes the hazard rate.

Proof. See Theorem 1.7.13 in Leadbetter et al. (1983). \square

Theorem 3.1.12 basically says that if the jump heights continue to be too large, there is no value u_n such that $F(u_n)$ is close to $1 - \tau/n$ and no non-degenerate limit distribution may be found. For discrete, integer-valued random variables X_1, \dots, X_n that are i.i.d. copies of X , with $x_F = \infty$, conditions (3.34) and (3.35) become

$$\lim_{k \rightarrow \infty} \frac{\bar{F}(k)}{\bar{F}(k-1)} = 1 \quad \text{and} \quad \lim_{k \rightarrow \infty} \frac{P(X=k)}{\bar{F}(k-1)} = 0,$$

respectively. If either of these conditions fails, we cannot find a non-degenerate limit distribution for $X_{(n)}$. In the following examples we will show that this is precisely the case for the Poisson and geometric distributions.

Example 3.1.13. (Poisson distribution) Let $X \sim \text{Poi}(\lambda)$, with $\lambda > 0$. Then

$$\begin{aligned} \frac{P(X=k)}{\bar{F}(k-1)} &= \frac{\lambda^k}{k!} \left(\sum_{l=k}^{\infty} \frac{\lambda^l}{l!} \right)^{-1} \\ &= \frac{\lambda^k}{k!} \left(\frac{\lambda^k}{k!} + \sum_{l=k+1}^{\infty} \frac{\lambda^l}{l!} \right)^{-1} \\ &= \left(1 + \sum_{l=k+1}^{\infty} \frac{k! \lambda^{l-k}}{l!} \right)^{-1}. \end{aligned}$$

The latter sum may be rewritten and estimated as

$$\sum_{s=1}^{\infty} \frac{\lambda^s}{(k+1) \cdot \dots \cdot (k+s)} \leq \sum_{s=1}^{\infty} \left(\frac{\lambda}{k} \right)^s = \frac{\lambda/k}{1 - \lambda/k} \quad \text{if } \lambda < k,$$

which tends to 0 as $k \rightarrow \infty$, and thus $P(X=k)/\bar{F}(k-1)$ tends to 1. Hence, by Theorem 3.1.12, no non-degenerate limit distribution exists for Poisson maxima, and $P[M_n \leq u_n] \rightarrow \rho$ only for $\rho = 0$ or 1.

Example 3.1.14. (Geometric distribution) Let $X \sim \text{Geo}(p)$, with the parameter $0 < p < 1$ denoting the success probability and the random variables X

counting the number of failures in 0-1 experiments before the first success, i.e. $P(X = k) = p(1 - p)^k$, for any $k \in \mathbb{Z}_+$. Then

$$\frac{P(X = k)}{\overline{F}(k - 1)} = \frac{p(1 - p)^k}{(1 - p)^k} = p \in (0, 1),$$

which violates condition (3.35).

We have now illustrated that for three of the most well-known and widely used discrete distributions, we may not find a limit law H as in Theorem 3.1.3. However, it should not be assumed that this problem occurs for discrete distributions in general. Feidt et al. (2010) give the following example of discrete distribution functions that do indeed possess a non-degenerate extreme value behaviour:

Example 3.1.15. Let $X \geq 0$ be an absolutely continuous random variable with distribution function F , probability density function f and right endpoint $x_F = \infty$, and suppose that its hazard rate $\lambda(y) = f(y)/\overline{F}(y) \rightarrow 0$, as $y \rightarrow \infty$. Denote by $\lceil X \rceil$ the integer-valued random variable with distribution function $\lceil F \rceil(y) = \lceil F \rceil(\lfloor y \rfloor) = F(\lfloor y \rfloor)$, for all $y \geq 0$, that is, let $\lceil X \rceil$ be a discretised version of X . Then, since \overline{F} is decreasing,

$$0 \leq \frac{P(\lceil X \rceil = k)}{\lceil \overline{F} \rceil(k - 1)} = \frac{F(k) - F(k - 1)}{\overline{F}(k - 1)} = \int_{k-1}^k \lambda(t) \frac{\overline{F}(t)}{\overline{F}(k - 1)} dt \leq \int_{k-1}^k \lambda(t) dt,$$

which tends to 0 as $k \rightarrow \infty$. Thus, $\lceil X \rceil$ satisfies condition (3.35). Moreover, note that since $\lceil F \rceil(y) \leq F(y)$ and $-\log \overline{F}(y) = \int_0^y \lambda(t) dt$ for all $y \geq 0$, we have

$$1 \leq \frac{\lceil \overline{F} \rceil(y)}{\overline{F}(y)} \leq \frac{\overline{F}(\lfloor y \rfloor)}{\overline{F}(\lfloor y \rfloor + 1)} = \exp \left(\int_{\lfloor y \rfloor}^{\lfloor y \rfloor + 1} \lambda(t) dt \right),$$

which, again by the above condition on the hazard rate, tends to 1 as $y \rightarrow \infty$. Thus, the two distribution functions F and $\lceil F \rceil$ are *tail-equivalent* (see, e.g., Definition 3.3.3 in Embrechts et al. (1997)). By Proposition 1.19 in Resnick (1987) it then follows that if F is in the maximum domain of attraction of an extreme value distribution H , then so is $\lceil F \rceil$, and vice versa. For instance, let X be Pareto distributed with parameters $\alpha, \phi > 0$ with distribution function F as in (3.6). Its hazard rate $\lambda(y) = \alpha/y$, for $y \geq \Phi$, vanishes as $y \rightarrow \infty$, as required. From Proposition 3.1.4 we know that $F \in \text{MDA}(\Phi_\alpha)$. By the above argument, we then also have that $\lceil F \rceil \in \text{MDA}(\Phi_\alpha)$.

3.1.3 Maxima of discrete random variables

In the previous subsection we discussed the non-existence of non-degenerate extremal limit laws for some well-known discrete distributions, like the binomial, geometric and Poisson. There is, however, a way to partially remedy this. Anderson et al. (1997) and Nadarajah and Mitov (2002) determined limit laws for precisely these distributions by allowing the distributional parameter (or one of them) to vary with the sample size n at suitable rates. This means that they actually considered triangular arrays,

$$\begin{array}{lll}
 m = 1 : & X_1^{(1)} & \sim \mathcal{P}_{\theta_1} \\
 m = 2 : & X_1^{(2)}, X_2^{(2)} & \stackrel{i.i.d.}{\sim} \mathcal{P}_{\theta_2} \\
 m = 3 : & X_1^{(3)}, X_2^{(3)}, X_3^{(3)} & \stackrel{i.i.d.}{\sim} \mathcal{P}_{\theta_3} \\
 \vdots & \vdots & \vdots \\
 m = n : & X_1^{(n)}, X_2^{(n)}, \dots, X_n^{(n)} & \stackrel{i.i.d.}{\sim} \mathcal{P}_{\theta_n},
 \end{array}$$

where \mathcal{P}_{θ_m} denotes a distribution with parameter, or collection of parameters, θ_m . To keep notations simple, we omit the superscript. Moreover, for any $y \in \mathbb{R}$, let $\lfloor y \rfloor$ denote the integer part of y , i.e. let $\lfloor y \rfloor = \max\{k \in \mathbb{Z} : k \leq y\}$, and similarly, let $\lceil y \rceil = \min\{k \in \mathbb{Z} : k \geq y\}$. For a discrete random variable X , taking values in \mathbb{Z} , we then naturally have $P(X = y) = P(X = \lfloor y \rfloor) = P(X = \lceil y \rceil)$ for all $y \in \mathbb{Z}$, whereas, for all $y \notin \mathbb{Z}$, $P(X = y) = 0$, and

$$\begin{aligned}
 P(X \leq y) &= P(X < y) = P(X \leq \lfloor y \rfloor) = P(X < \lceil y \rceil), \\
 P(X > y) &= P(X \geq y) = P(X > \lfloor y \rfloor) = P(X \geq \lceil y \rceil).
 \end{aligned} \tag{3.36}$$

Having clarified notations, we now give the extremal limit results for the Poisson, binomial and geometric distributions, the first of which was proven by Anderson et al. (1997) and the latter two by Nadarajah and Mitov (2002). Proposition 3.1.16 first uses convergence of the Poisson distribution with large parameter λ to the normal distribution and then the fact that the normal distribution is in the maximum domain of attraction of the Gumbel distribution, as shown in Proposition 3.1.4(d). Proposition 3.1.17 proceeds similarly for the binomial distribution.

Proposition 3.1.16. *(Poisson distribution, Anderson et al., 1997) For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. Poisson random variables with parameter*

$\lambda = \lambda_n > 0$. If λ_n grows with n such that $(\log n)^3 = o(\lambda_n)$, then, for all $x \in \mathbb{R}$,

$$P\left(X_{(n)} \leq \sqrt{\lambda_n} \alpha_n x + \lambda_n + \sqrt{\lambda_n} \beta_n\right) \rightarrow e^{-e^{-x}} = \Lambda(x),$$

as $n \rightarrow \infty$, where

$$\alpha_n = \frac{1}{\sqrt{2 \log n}}, \quad \beta_n = \sqrt{2 \log n} - \frac{\log \log n + \log 4\pi}{2\sqrt{2 \log n}}.$$

□

Proposition 3.1.17. (*Binomial distribution, Nadarajah and Mitov, 2002*) For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. binomial random variables with number of trials $N = N_n \in \{1, 2, \dots\}$ and fixed probability of success $p \in (0, 1)$. If $N_n \rightarrow \infty$ as $n \rightarrow \infty$ such that $(\log n)^3 = o(N_n)$, then, for all $x \in \mathbb{R}$,

$$P\left(X_{(n)} \leq \sqrt{p(1-p)N_n} \alpha_n x + pN_n + \sqrt{p(1-p)N_n} \beta_n\right) \rightarrow e^{-e^{-x}} = \Lambda(x),$$

as $n \rightarrow \infty$, where

$$\alpha_n = \frac{1}{\sqrt{2 \log n}}, \quad \beta_n = \sqrt{2 \log n} - \frac{\log \log n + \log 4\pi}{2\sqrt{2 \log n}}.$$

□

Note that the geometric distribution treated by Nadarajah and Mitov (2002) as a special case of the negative binomial is the shifted geometric distribution, which counts the number of trials until the first success. In contrast, the following proposition treats the geometric distribution which counts the number of failures before the first success. The normalising constants however remain the same as those used by Nadarajah and Mitov (2002).

Proposition 3.1.18. (*Geometric distribution*) For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. geometric random variables with probability of success $p = p_n \in (0, 1)$, probability mass function $P(X_1 = k) = p(1-p)^k$ and cumulative distribution function $F(k) = 1 - (1-p)^{k+1}$, for any $k \in \mathbb{Z}_+$. If $p_n \rightarrow 0$ as $n \rightarrow \infty$, then, for all $x \in \mathbb{R}$,

$$P\left(X_{(n)} \leq \frac{\log n + x}{p_n}\right) \rightarrow e^{-e^{-x}} = \Lambda(x), \quad \text{as } n \rightarrow \infty.$$

Proof. Fix any $x \in \mathbb{R}$. Due to (3.36) and (3.7), we have

$$\begin{aligned} & P\left(X_{(n)} \leq \frac{\log n + x}{p_n}\right) \\ &= P\left(X_{(n)} \leq \left\lfloor \frac{\log n + x}{p_n} \right\rfloor\right) = F^n\left(\left\lfloor \frac{\log n + x}{p_n} \right\rfloor\right) \\ &= \left\{1 - (1 - p_n)^{\left\lfloor \frac{\log n + x}{p_n} \right\rfloor + 1}\right\}^n = \left\{1 - e^{(\left\lfloor \frac{\log n + x}{p_n} \right\rfloor + 1) \log(1 - p_n)}\right\}^n, \end{aligned}$$

for any integer $n \geq e^{-x}$. Using $z - 1 \leq \lfloor z \rfloor \leq z$, for $z \in \mathbb{R}$, and $-p/(1 - p) \leq \log(1 - p) \leq -p$, for all $p \in (0, 1)$, we obtain

$$\left\{1 - e^{-(\log n + x)}\right\}^n \leq P\left(X_{(n)} \leq \frac{\log n + x}{p_n}\right) \leq \left\{1 - e^{-\frac{\log n + x}{1 - p_n}} \cdot (1 - p_n)\right\}^n. \quad (3.37)$$

As $n \rightarrow \infty$ and $p_n \rightarrow 0$, both sides of (3.37) tend to $\exp\{-e^{-x}\} = \Lambda(x)$. \square

The following proposition investigates the rate of convergence of the limit result from Proposition 3.1.18 and suggests two improvements. One way to reduce the error is to approximate by a discretised version of the Gumbel distribution, the other is to use different normalising constants.

Proposition 3.1.19. (*Geometric distribution*) For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. geometric random variables with success probability $p_n \in (0, 1)$, failure probability $q_n = 1 - p_n$, probability mass function $P(X_1 = k) = p_n q_n^k$ and survival function $\bar{F}(k) = q_n^{k+1}$, for any $k \in \mathbb{Z}_+$. Then:

(a) (*Approximation by a discretised Gumbel distribution*) For all $k \in \mathbb{Z}_+$ and for all $k^* \in \mathbb{R}$ defined by $k^* = -\log n + k \log(1/q_n)$,

$$\left|P\left(X_{(n)} < \frac{\log n + k^*}{\log(1/q_n)}\right) - e^{-e^{-k^*}}\right| \leq \frac{\log n}{q_n n} + \frac{1}{n} =: \delta_{\text{PoiAppr}}. \quad (3.38)$$

(b) (*Approximation by a Gumbel distribution*) For all $x \in \mathbb{R}$,

$$\left|P\left(X_{(n)} < \frac{\log n + x}{\log(1/q_n)}\right) - e^{-e^{-x}}\right| \leq \delta_{\text{PoiAppr}} + e^{-1} \log(1/q_n) =: \delta_{\text{Cont}}.$$

(c) (*Using the normalising constants from Nadarajah and Mitov, 2002*)

$$\left|P\left(X_{(n)} < \frac{\log n + x}{1 - q_n}\right) - e^{-e^{-x}}\right| \leq \delta_{\text{Cont}} + \frac{1 - q_n}{2q_n} (\log^2 n + e^{-1}).$$

Note that the failure probability q_n need not vary with the sample size n for approximation by a discretised Gumbel distribution. The error bound is sharp for any constant $q_n \equiv q \in (0, 1)$, showing clearly that it makes more sense to approximate a discrete distribution by another discrete distribution than by a continuous one, as there is no need to add an extra error as in (b).

The extra error in (b), $e^{-1} \log(1/q_n)$, is the discretisation error that arises when going from the Gumbel concentrated on the lattice of points k^\star to the continuous Gumbel distribution over \mathbb{R} . It dominates the overall error in (b) unless q_n tends to 1 fast enough as $n \rightarrow \infty$, that is, unless $1 - q_n = O(\log(n)/n)$, in which case the discretisation error is of the same order as the first error term from (a).

Part (c) shows that the choice of normalising constants, more precisely, of the scaling by p_n in Proposition 3.1.18, is far from optimal. In order for the approximation in (c) to be good we require $p_n = o(1/\log^2 n)$. Its being a stronger condition than the one for the asymptotic result from Proposition 3.1.18 is justified by (c) also being a stronger result in the sense that it gives a uniform bound. The error in (c) is of the same order as the error in (a) only if $1 - q_n = O(1/(n \log n))$.

Proof. For ease of notation we omit the subscript n . (a) Let $A = [y, \infty)$ for any choice of $y \geq 0$. Then, by (3.36), $P(X_1 \in A) = q^{\lceil y \rceil}$, and, setting $k := \lceil y \rceil \in \mathbb{Z}_+$, Theorem 3.1.1 gives

$$\left| P(X_{(n)} < k) - e^{-nq^k} \right| \leq q^k. \quad (3.39)$$

With $k^\star \in \mathbb{R}$ chosen such that $k = (\log n + k^\star)/\log(1/q)$, we then have

$$\left| P\left(X_{(n)} < \frac{\log n + k^\star}{\log(1/q)}\right) - e^{-e^{-k^\star}} \right| \leq \frac{e^{-k^\star}}{n}. \quad (3.40)$$

In order to find a uniform bound for all $k \in \mathbb{Z}_+$, choose $x_0 := x_{0n} := -\log \log n$. Then, for all k such that $k^\star \geq x_0$, we have $\exp(-k^\star)/n \leq \exp(-x_0)/n = \log(n)/n$, whereas for k such that $k^\star \leq x_0$, we may bound the error in (3.40) by further adding the Gumbel distribution to the error at m^\star , where $m := \lfloor y_0 \rfloor := \lfloor (\log n + x_0)/\log(1/q) \rfloor$, i.e.

$$\begin{aligned} \left| P\left(X_{(n)} < \frac{\log n + k^\star}{\log(1/q)}\right) - e^{-e^{-k^\star}} \right| &\leq \frac{e^{-m^\star}}{n} + e^{-e^{-m^\star}} = e^{-m \log(1/q)} + e^{-e^{-m^\star}} \\ &\leq \frac{e^{-x_0}}{qn} + e^{-e^{-x_0}} \leq \frac{\log n}{qn} + \frac{1}{n}, \end{aligned} \quad (3.41)$$

where we used $m \geq y_0 - 1$ in the second inequality. See Figure 3.2 for a sketch. (3.41) provides a bound for all k^* , and thus

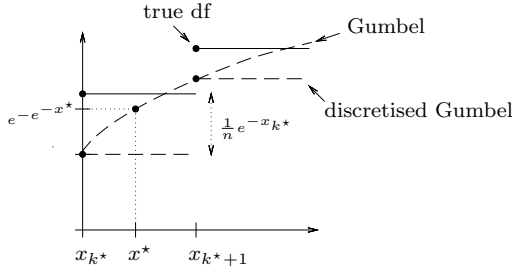


Figure 3.2: At all values $k^* \leq m^* \leq x_0$, the error cannot exceed the sum of the bound e^{-m^*}/n on the difference between the two distributions functions and the height $\exp\{-e^{-x_0}\}$ of the Gumbel distribution.

$$\left| P\left(X_{(n)} < \frac{\log n + k^*}{\log(1/q)}\right) - e^{-e^{-k^*}} \right| \leq \frac{\log n}{qn} + \frac{1}{n}.$$

(b) Let $x = \log(1/q)y - \log n$. By adding and subtracting $\exp\{-e^{-x}\}$ into (3.38), and noting that, since $y \leq \lceil y \rceil = k$, we have $x \leq k^*$ and $\exp\{-e^{-k^*}\} - \exp\{-e^{-x}\} \geq 0$. We thus obtain

$$\left| P\left(X_{(n)} < \frac{\log n + x}{\log(1/q)}\right) - e^{-e^{-x}} \right| \leq \delta_{\text{PoiAppr}} + e^{-e^{-k^*}} - e^{-e^{-x}},$$

where

$$\begin{aligned} e^{-e^{-k^*}} - e^{-e^{-x}} &\leq \int_x^{k^*} e^{-t} e^{-e^{-t}} dt \leq e^{-1}(k^* - x) \\ &= e^{-1} \log(1/q)(\lceil y \rceil - y) \leq e^{-1} \log(1/q). \end{aligned}$$

(c) From (a) and (b) we have

$$\left| P\left(X_{(n)} < \frac{\log n + x}{\log(1/q)}\right) - e^{-e^{-x}} \right| \leq \frac{e^{-x}}{n} + e^{-1} \log(1/q). \quad (3.42)$$

Choose $x' \geq -\log n$ such that

$$y = \frac{\log n + x}{\log(1/q)} = \frac{\log n + x'}{1 - q}.$$

By adding and subtracting $\exp\{-e^{-x'}\}$ into (3.42) and observing that $x > x'$ since $\log(1/q) > 1 - q$, we then obtain

$$\left| P\left(X_{(n)} < \frac{\log n + x'}{1 - q}\right) - e^{-e^{-x'}} \right| \leq \frac{e^{-x'}}{n} + e^{-1} \log(1/q) + e^{-e^{-x}} - e^{-e^{-x'}}.$$

For the latter error term we find

$$\begin{aligned} e^{-e^{-x}} - e^{-e^{-x'}} &= e^{-e^{-x}} \left[1 - e^{-(e^{-x'} - e^{-x})} \right] \leq e^{-x'} \left[1 - e^{-(x - x')} \right] \\ &\leq e^{-x'} (x - x') = e^{-x'} [\log(1/q) - (1 - q)] y, \end{aligned}$$

where we used $\exp\{-e^{-x}\} \leq 1$ in the first inequality and $1 - e^{-z} \leq z$ for $z \geq 0$ in both inequalities. Note that use of the definition of the logarithm and the geometric series give

$$\begin{aligned} &\log(1/q) - (1 - q) \\ &= \sum_{j=2}^{\infty} \frac{(1 - q)^j}{j} \leq \sum_{j=2}^{\infty} \frac{(1 - q)^j}{2} = \frac{1}{2} \left[\sum_{j=0}^{\infty} (1 - q)^j - 1 - (1 - q) \right] \\ &= \frac{1}{2} \left[\frac{1}{q} - 2 + q \right] = \frac{(1 - q)^2}{2q}. \end{aligned}$$

Then,

$$\begin{aligned} e^{-e^{-x}} - e^{-e^{-x'}} &\leq \frac{(1 - q)^2 y}{2q} e^{-x'} = \frac{1 - q}{2q} (\log n + x') e^{-x'} \\ &\leq \frac{1 - q}{2q} (e^{-x'} \log n + e^{-1}). \end{aligned}$$

Thus,

$$\begin{aligned} &\left| P\left(X_{(n)} < \frac{\log n + x'}{1 - q}\right) - e^{-e^{-x'}} \right| \\ &\leq \frac{e^{-x'}}{n} + e^{-1} \log(1/q) + \frac{1 - q}{2q} (e^{-x'} \log n + e^{-1}). \end{aligned}$$

For a uniform bound over all x' , we choose x_0 as before in (a) and (b), and obtain, with an analogous argument, the overall error bound

$$\frac{\log n}{n} + e^{-1} \log(1/q) + \frac{1 - q}{2q} (\log^2 n + e^{-1}) + \frac{1}{n}.$$

□

3.2 Poisson process approximation for MPPE's

Theorem 3.1.1 not only gives information on distributional approximation for the maximum $X_{(n)}$ of i.i.d. random variables X_1, \dots, X_n with distribution function F , as in (3.2), but also on approximation of the laws of each of the order statistics of this sample. Denote by $X_{(k)}$ the k -th order statistic of the sample X_1, \dots, X_n , i.e. order the sample as follows:

$$\min_{1 \leq i \leq n} X_i = X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} = \max_{1 \leq i \leq n} X_i.$$

The number of points exceeding a threshold u_n can be related to each order statistic by

$$\left\{ \sum_{i=1}^n I_{\{X_i > u_n\}} \leq k \right\} = \{X_{(n-k)} \leq u_n\},$$

and Theorem 3.1.1 gives

$$\left| P(X_{(n-k)} \leq u_n) - e^{-n\bar{F}(u_n)} \frac{(n\bar{F}(u_n))^k}{k!} \right| \leq \bar{F}(u_n).$$

Several further generalisations can be achieved by using point processes. One of them incorporates a way to specify which of the X_i 's are the ones exceeding the threshold. This is not immediately given by Theorem 3.1.1. The object that we are studying, $\sum_{i=1}^n I_{\{X_i > u_n\}}$, needs to be generalised so as to additionally pinpoint the indices of the random variables exceeding u_n . This can be attained by using *point processes of exceedances* (PPE's). Classically, a PPE is a point process of the form

$$N_n = \sum_{i=1}^n I_{\{X_i > u\}} \delta_{in^{-1}}, \quad (3.43)$$

that lives on the state space $E = (0, 1]$. If applied to the entire state space, this point process recovers the total number of extreme points, i.e.

$$N_n((0, 1]) = \sum_{i=1}^n I_{\{X_i > u\}},$$

but if applied to a measurable subset $B \subset (0, 1]$, $N_n(B)$ gives only the random number of X_i 's that exceed u_n and for which $in^{-1} \in B$. For instance, suppose that the X_i 's describe the outcomes of n identical and independent random

experiments that are realised at consecutive time points i . The random number of extreme points that occur after time $t > 0$ is then given by $N_n((tn^{-1}, 1])$. Now what do we know about the distribution of N_n ? Embrechts et al. (1997) on p.238 states the following theorem for weak convergence of point processes of exceedances to a Poisson process:

Theorem 3.2.1. *Suppose that (X_n) is a sequence of i.i.d. random variables with common distribution function F . Let (u_n) be threshold values such that for some $\tau \in (0, \infty)$,*

$$n\bar{F}(u_n) = \mathbb{E} \sum_{i=1}^n I_{\{X_i > u_n\}} \rightarrow \tau, \text{ for } n \rightarrow \infty.$$

Then the point processes of exceedances N_n , as defined in (3.43), converge weakly in $M_p(E)$ to a homogeneous Poisson process N on $E = (0, 1]$ with intensity τ , i.e. N is $\text{PRM}(\tau|\cdot|)$, where $|\cdot|$ denotes Lebesgue measure on E .

Another kind of point processes studied in EVT are *marked point processes* (MPP's) of the form $\sum_{i=1}^n \delta_{X_i}$, that live on the state space E of the random variables X_i . An MPP gives a random configuration of points in space and counts the number of points in any measurable subset of the state space that it is applied to. The MPP converges weakly in $M_p(E)$ to a Poisson process with mean measure λ if and only if its mean measure converges vaguely to λ , as $n \rightarrow \infty$ (see Chapter 3 in Resnick (1987) for more details).

We introduce yet another kind of point process, that we call *marked point process of exceedances* (MPPE) and that we define as follows:

$$\Xi_{u,n} := \sum_{i=1}^n I_{\{X_i > u\}} \delta_{X_i}. \quad (3.44)$$

Though the MPPE does not mark the points that exceed u_n as the PPE does, it contains more information relevant to the study of extreme values than an MPP, as it is not only a random configuration of points in space, but specifically a random configuration of points exceeding a threshold. From now on, we concentrate on MPPE's, as they are better suited to our purposes than PPE's or MPP's.

The state space $E \subseteq \mathbb{R}$ of $\Xi_{u,n}$ is the set of values $X(\Omega)$ taken by the i.i.d. random variables X_1, \dots, X_n . Let $\mathcal{E} = \mathcal{B}(E)$. For a fixed set $B \in \mathcal{E}$, $\Xi_{u,n}(B)$ gives the random number of X_i 's in $(u, \infty) \cap B$, whereas for a fixed $\omega \in \Omega$,

$\Xi_{u,n}(\omega) = \sum_{i \in \{1, \dots, n\}: x_i > u} \delta_{x_i}$ gives the point configuration of the realisations $X_i(\omega) = x_i > u$. We can generalise MPPE's by considering a set $A \in \mathcal{E}$ of an arbitrary shape instead of a threshold u :

$$\Xi_A := \Xi_{A,n} := \sum_{i=1}^n I_{\{X_i \in A\}} \delta_{X_i}. \quad (3.45)$$

A should be chosen such that points x_i lying in it can be considered to be extreme points. This generalisation makes sense mostly for higher-dimensional points, i.e. points in \mathbb{R}^d , $d \geq 2$, as it gives more flexibility to the choice of region in which it makes sense to consider points to be extreme. For instance, to obtain a multivariate analogue to (3.44) with componentwise thresholds, we can simply choose $A = (u_1, \infty) \times \dots \times (u_d, \infty)$, for $u_1, \dots, u_d \in \mathbb{R}$. Another example would be to choose A as the complement of a disc with a 'large' radius r centred in some origin. The mean measure of Ξ_A on \mathcal{E} is given by

$$\lambda(\cdot) := \lambda_A(\cdot) := \mathbb{E} \Xi_A(\cdot) = \sum_{i=1}^n P(X_i \in A \cap \cdot) = nP(X \in A \cap \cdot).$$

Note that Ξ_A is a finite point process and that λ is a finite measure.

As both PPE's and MPP's asymptotically, under certain conditions, behave like Poisson processes, the same can be expected of the MPPE's. We can indeed readily apply Theorem 2.5.13 from Section 2.5.5 to MPPE's to obtain a result in this vein. Our result, however, is stronger than a mere limit theorem: Proposition 3.2.2 below gives an estimate of the error of the approximation, in the total variation distance, of the law of Ξ_A by that of a Poisson process with mean measure $\mathbb{E} \Xi_A$.

Proposition 3.2.2. *For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. copies of an E -valued random variable X . For a fixed set $A \in \mathcal{E}$, let $\Xi_A = \sum_{i=1}^n I_{\{X_i \in A\}} \delta_{X_i}$ be the marked point process of points in A and let $W_A = \sum_{i=1}^n I_{\{X_i \in A\}}$ denote the random number of points in A . Then,*

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E} \Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E} W_A)) \leq P(X \in A). \quad (3.46)$$

Proof. Let $P_A = \mathcal{L}(X|X \in A)$ and define an i.i.d. random sample X'_1, \dots, X'_n with common distribution P_A that is independent of the sample X_1, \dots, X_n . Then the process $\sum_{i=1}^n I_{\{X_i \in A\}} \delta_{X'_i}$ has the same distribution as the process of interest Ξ_A . Note that due to the independence of the samples X_1, \dots, X_n and

X'_1, \dots, X'_n , the process $\sum_{i=1}^n I_{\{X_i \in A\}} \delta_{X'_i}$ is distributed as $\sum_{j=1}^{W_A} \delta_{Z'_j}$, where the Z'_j are independent, have common distribution P_A , and are independent of W_A . Furthermore, note that a $\text{PRM}(\mathbb{E}\Xi_A)$ can be realised as $\sum_{j=1}^{W^*} \delta_{Z'_j}$, where $W^* \sim \text{Poi}(\mathbb{E}W_A)$ is independent of the Z'_j . It then follows from the proof of Theorem 2.5.13 that

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)),$$

where the latter is at most $P(X \in A)$ by Theorem 3.1.1. \square

The error of Poisson process approximation for an MPPE is thus the same as the error for Poisson approximation for the number of extreme points. This means that we may (more or less) recover the results from Propositions 3.1.5 - 3.1.9 and 3.1.19. However, in order to do this, we first need to subject the random variables X_i to a normalisation as we did in these propositions. More precisely, define the normalised random variable

$$X^* := \frac{X - b_n}{a_n}, \quad (3.47)$$

where $a_n, b_n \in \mathbb{R}$ with $a_n > 0$, and let X_1^*, \dots, X_n^* be i.i.d. copies of X^* , with state space $E^* = X^*(\Omega)$. Let $\mathcal{E}^* = \mathcal{B}(E^*)$ and let A^* be the accordingly normalised version of A , i.e. $A^* = ((A - b_n)/a_n) \in \mathcal{E}^*$. For instance, for $E = [0, \infty)$ and $A = [u, \infty)$ for $u \geq 0$, we have $E^* = [-b_n/a_n, \infty)$ and $A^* = [u^*, \infty)$, where $u^* = (u - b_n)/a_n \geq -b_n/a_n$. The distribution function of X^* is given by

$$F^*(x) = P(X^* \leq x) = P(X \leq a_n x + b_n) = F(a_n x + b_n), \quad x \in E^*. \quad (3.48)$$

For any $A \in E$ and $A^* \in E^*$, we thus have

$$P(X \in A) = P(X^* \in A^*) \quad \text{and} \quad W_A = \sum_{i=1}^n I_{\{X_i^* \in A^*\}} =: W_{A^*}^*. \quad (3.49)$$

Similarly, we have $\Xi_A = \sum_{i=1}^n I_{\{X_i \in A\}} \delta_{X_i}$, which still lives on the state space E . We define a normalised version of this process,

$$\Xi_{A^*}^* := \sum_{i=1}^n I_{\{X_i^* \in A^*\}} \delta_{X_i^*}, \quad (3.50)$$

which has state space E^* and mean measure

$$\lambda^*(.) := \mathbb{E}\Xi_{A^*}^*(.) = nP(X^* \in A^* \cap .), \quad (3.51)$$

on \mathcal{E}^* . Obviously, $\Xi_A \neq \Xi_{A^*}^*$. However, for any $R \in \mathcal{M}_p(E)$ and its normalised version $R^* \in \mathcal{M}_p(E^*)$, we have $P(\Xi_A \in R) = P(\Xi_{A^*}^* \in R^*)$. It follows that

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\lambda)) = d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda^*)), \quad (3.52)$$

and (3.46) may equivalently be expressed as

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq d_{TV}(\mathcal{L}(W_{A^*}^*), \text{Poi}(nP(X^* \in A^*))) \leq P(X^* \in A^*). \quad (3.53)$$

We now use (3.53) to obtain process analogues of Propositions 3.1.5 - 3.1.9 and 3.1.19, that is, we determine error bounds for Poisson process approximation of MPPE's with marks X_1, \dots, X_n following well-known distributions. Section 3.2.1 discusses the case of continuous marks and treats all distributions listed in Proposition 3.1.4. In Section 3.2.2 we discuss an example of an MPPE with discrete marks. More precisely, we suppose that X_1, \dots, X_n follow the geometric distribution. We first approximate the MPPE with geometric marks by a Poisson process on \mathbb{Z}_+ . Then, as processes with a continuous intensity are more practicable, we further try to approximate the MPPE by a Poisson process with continuous intensity over \mathbb{R}_+ . To achieve this, we use the weaker d_2 -metric and the results from Section 2.6. Throughout Section 3.2, we set, for simplicity, very strong assumptions on our MPPE's, as we require i.i.d. marks and i.i.d. indicators. Our main efforts therefore lie, not so much in determining an error bound for basic Poisson approximation, but rather in determining error bounds between two Poisson processes, if necessary. Section 3.2.3 gives a short discussion on what might happen for different assumptions.

3.2.1 Application to MPPE's with continuous marks

Suppose that the distribution function F of the i.i.d. E -valued random variables X, X_1, \dots, X_n is absolutely continuous with probability density function $f(y)$, $y \in E$. We may then define an intensity function $\lambda(y) := nf(y)$ for all $y \in E$, and write the intensity measure λ of Ξ_A as

$$\lambda(B) = \int_{A \cap B} \lambda(y) dy, \quad \text{for any } B \in \mathcal{E}.$$

Due to (3.48), the probability density function of X^* is given by

$$f^*(x) = \frac{d}{dx} F(a_n x + b_n) = a_n f(a_n x + b_n), \quad (3.54)$$

and we may write the intensity measure λ^* of $\Xi_{A^*}^*$ as

$$\lambda^*(B^*) = \int_{A^* \cap B^*} \lambda^*(x) dx, \quad \text{for any } B^* \in \mathcal{E}^*, \quad (3.55)$$

where $\lambda^*(x) := n f^*(x)$. Propositions 3.2.3 - 3.2.7 below determine, on the one hand, the intensity functions λ^* of MPPE's whose marks follow well-known continuous distributions, and, on the other hand, the errors in total variation arising when approximating these MPPE's by a Poisson process with intensity function λ^* . Knowledge of the intensity function λ^* simplifies the computation of $\lambda^*(B^*) = n P(X^* \in B^*)$ for measurable sets $B^* \subseteq E^*$ of arbitrary shape, as well as of the error $P(X^* \in A^*) = n^{-1} \lambda^*(A^*)$ that we obtain from Proposition 3.2.2, or, equivalently, from (3.53). For each of the propositions below, we choose $A^* = [u^*, x_{F^*}^*] \cap E^*$. Then, the smaller the choice of the threshold u^* , the bigger the error $P(X^* \in A^*) = P(X^* \geq u^*)$, and the worse the approximation by a Poisson process. This of course exactly mirrors Poisson approximation for the number of points exceeding u^* , for which we have the same error $P(X^* \geq u^*)$. Put in another way, the distribution of the number of points exceeding u^* is binomial with success probability precisely equal to $P(X^* \geq u^*)$. The smaller the threshold, the bigger the success probability, and the worse Poisson approximation which requires a success probability tending to zero. We thus aim for a high threshold u^* and consequently a small number of exceedances. For each of the propositions below we will discuss suitable choices of u^* depending on the sample size n .

Proposition 3.2.3. (*Exponential distribution*) *For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. exponential random variables with parameter $\lambda > 0$. Define the normalised random variables $X_i^* = \lambda X_i - \log n$, $i = 1, \dots, n$, taking values in $E^* = [-\log n, \infty)$. Let $A^* = [u^*, \infty)$ for any choice of $u^* \in E^*$, and let $W_{A^*}^*$ and $\Xi_{A^*}^*$ be defined as in (3.49) and (3.50), respectively. Then the mean measure of $\Xi_{A^*}^*$ is given by*

$$\lambda^*(B^*) = \int_{A^* \cap B^*} e^{-x} dx, \quad \text{for any } B^* \in \mathcal{E}^*,$$

and

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq d_{TV}(\mathcal{L}(W_{A^*}^*), \text{Poi}(e^{-u^*})) \leq \frac{e^{-u^*}}{n}.$$

Proof. We have $f(y) = \lambda e^{-\lambda y}$ for all $y \geq 0$, $a_n = \lambda^{-1}$ and $b_n = \lambda^{-1} \log n$. By (3.54) and (3.55), $f^*(x) = e^{-x}/n$ and $\lambda^*(x) = e^{-x}$ for all $x \geq -\log n$, and $\lambda^*(B^*) = \int_{A^* \cap B^*} \lambda^*(x) dx$ for any $B^* \in \mathcal{B}([-\log n, \infty))$. By (3.53),

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq P(X_1^* \geq u^*) = \frac{e^{-u^*}}{n}.$$

□

The expected number of exceedances of the threshold $u^* = u_n^*$ is e^{-u^*} , whereas the error of the approximation in the total variation distance is e^{-u^*}/n . Thus, the lower we set the threshold u^* with respect to the sample size, the more exceedances we will expect and the bigger the error of the approximation by a Poisson process will be. For instance, for $u_n^* = -\log \log n$, the error estimate is $\log(n)/n$ and we expect about $\log n$ exceedances. Round $\log n$ to its nearest integer value $[\log n]$. The points of the point process $\Xi_{A^*}^*$ are then (more or less) the realisations of the $[\log n]$ biggest normalised order statistics $X_{(n-[\log n]+1)}^*, \dots, X_{(n-1)}^*, X_{(n)}^*$ of the sample. For $u_n^* \equiv 0$ on the other hand, we expect only 1 threshold exceedance among n random variables and the single expected point of $\Xi_{A^*}^*$ is thus the realisation of the maximum $X_{(n)}$ of the sample. The error estimate is of size $1/n$ and thereby decreases more rapidly as the sample size n increases.

Proposition 3.2.4. (*Pareto distribution*) For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. Pareto random variables with shape and scale parameters $\alpha, \phi > 0$. Define the normalised random variables $X_i^* = \phi^{-1} n^{-1/\alpha} X_i$, $i = 1, \dots, n$, taking values in $E^* = [n^{-1/\alpha}, \infty)$. Let $A^* = [u^*, \infty)$ for any choice of $u^* \in E^*$, and let $W_{A^*}^*$ and $\Xi_{A^*}^*$ be defined as in (3.49) and (3.50), respectively. Then the mean measure of $\Xi_{A^*}^*$ is given by

$$\lambda^*(B^*) = \int_{A^* \cap B^*} \alpha x^{-\alpha-1} dx, \quad \text{for any } B^* \in \mathcal{E}^*,$$

and

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq d_{TV}\left(\mathcal{L}(W_{A^*}^*), \text{Poi}\left(\frac{1}{u^{*\alpha}}\right)\right) \leq \frac{1}{nu^{*\alpha}}.$$

Proof. We have $f(y) = \alpha\phi^\alpha y^{-\alpha-1}$ for all $y \geq \phi$, $a_n = \phi n^{1/\alpha}$ and $b_n \equiv 0$. By (3.54) and (3.55), $f^*(x) = \alpha n^{-1} x^{-\alpha-1}$ and $\lambda^*(x) = \alpha x^{-\alpha-1}$ for all $x \geq n^{-1/\alpha}$, and $\lambda^*(B^*) = \int_{A^* \cap B^*} \lambda^*(x) dx$ for any $B^* \in \mathcal{B}([n^{-1/\alpha}, \infty))$. By (3.53),

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq P(X_1^* \geq u^*) = \frac{1}{nu^{\star\alpha}}.$$

□

For roughly $\log n$ expected threshold exceedances among the n i.i.d. Pareto random variables, and an error of order $\log(n)/n$, we need to choose $u^* = u_n^* = (\log n)^{-1/\alpha}$. Similarly, for only 1 threshold exceedance (by the maximum of the Pareto variables) and a rather smaller error $1/n$, we would have to set $u_n^* \equiv 1$.

Proposition 3.2.5. (*Uniform distribution*) For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. uniform random variables with parameters $a, b \in \mathbb{R}$ such that $a < b$. Define the normalised random variables $X_i^* = -n(b - X_i)/(b - a)$, $i = 1, \dots, n$, taking values in $E^* = [-n, 0)$. Let $A^* = [u^*, 0)$ for any choice of $u^* \in E^*$, and let $W_{A^*}^*$ and $\Xi_{A^*}^*$ be defined as in (3.49) and (3.50), respectively. Then the mean measure of $\Xi_{A^*}^*$ is given by

$$\lambda^*(B^*) = \int_{A^* \cap B^*} 1 dx, \quad \text{for any } B^* \in \mathcal{E}^*,$$

and

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq d_{TV}(\mathcal{L}(W_{A^*}^*), \text{Poi}(-u^*)) \leq \frac{-u^*}{n}.$$

Proof. We have $f(y) = 1/(b - a)$ for all $y \in [a, b)$, and we use the normalisation $y = x(b - a)/n + b$, where $x \in [-n, 0)$. Then

$$f^*(x) = \frac{d}{dx} F\left(\frac{b-a}{n}x + b\right) = f\left(\frac{b-a}{n}x + b\right) \frac{b-a}{n} = \frac{1}{n}$$

and $\lambda^*(x) = 1$, for all $x \in [-n, 0)$. Moreover, $\lambda^*(B^*) = \int_{A^* \cap B^*} \lambda^*(x) dx$ for any $B^* \in \mathcal{B}([-n, 0))$. By (3.53),

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq d_{TV}(\mathcal{L}(W_{A^*}^*), \text{Poi}(-u^*)) \leq P(X_1^* \geq u^*) = \frac{-u^*}{n}.$$

□

Here, e.g. $u_n^* = -\sqrt{n}$ will lead to $\Xi_{A^*}^*$ capturing roughly the \sqrt{n} biggest order statistics and an error estimate of size $1/\sqrt{n}$, whereas the choice $u_n^* = -\log \log n$ will give $\log n$ expected exceedances and the error estimate $\log(n)/n$.

The intensity function of the MPPE with i.i.d. normal marks is given by $\lambda_n^*(x) = na_n\varphi(a_nx + b_n)$, where a_n, b_n are the norming constants defined in (3.12). As seen in Proposition 3.1.8 (c), this intensity function may be approximated by e^{-x} as $n \rightarrow \infty$. With $A^* = [u^*, \infty)$, we then expect about e^{-u^*} threshold exceedances. Proposition 3.2.6 below gives estimates for the two steps involved in approximating $\mathcal{L}(\Xi_{A^*}^*)$ by $\text{PRM}(\int_{A^* \cap \cdot} e^{-x} dx)$. The (much) bigger of the two estimates is the one that arises from approximating $\text{PRM}(\int_{A^* \cap \cdot} \lambda_n^*(x))$ by $\text{PRM}(\int_{A^* \cap \cdot} e^{-x} dx)$. We would like to get a positive, but not too big, expected number of threshold exceedances. The choice $u^* = -\log \log n$, which gives an expected number of $\log n$ threshold exceedances, is not suitable in view of error bound (b) of Proposition 3.2.6, which becomes too big. Instead, better choose $u^* = -\alpha \log \log n$ for any $\alpha \in (0, 1)$. We then expect about $(\log n)^\alpha$ threshold exceedances and obtain an error estimate of order $(\log \log n)^2 / (\log n)^{1-\alpha}$.

Proposition 3.2.6. *(Standard normal distribution) For each integer $n \geq 5$, let X_1, \dots, X_n be i.i.d. standard normal random variables, with probability density function $\varphi(y) = (2\pi)^{-1} e^{-y^2/2}$ for all $y \in \mathbb{R}$. Define the normalised random variables $X_i^* = a_n^{-1}(X_i - b_n)$, $i = 1, \dots, n$, with a_n and b_n from (3.12) and with state space $E^* = \mathbb{R}$. Let $A^* = [u^*, \infty)$ for any choice of $u^* \in [-\alpha \log \log n, 0]$, where $\alpha \in (0, 1)$, and let $\Xi_{A^*}^*$ be defined as in (3.50). Then the mean measure of $\Xi_{A^*}^*$ is given by*

$$\lambda_n^*(B^*) = \int_{A^* \cap B^*} na_n\varphi(a_nx + b_n)dx, \quad \text{for any } B^* \in \mathcal{E}^*,$$

Moreover, define

$$\lambda^*(B^*) = \int_{A^* \cap B^*} e^{-x}dx, \quad \text{for any } B^* \in \mathcal{E}^*.$$

Then

- (a) $d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda_n^*)) \leq \frac{6e^{-u^*}}{n},$
- (b) $d_{TV}(\text{PRM}(\lambda_n^*), \text{PRM}(\lambda^*)) \leq \frac{(3 \log \log n + \log 4\pi)^2}{16 \log n} e^{-u^*}.$

Proof. (a) By (3.54) and (3.55), $\lambda_n^*(x) = na_n\varphi(a_nx + b_n)$ for all $x \in \mathbb{R}$, and $\lambda_n^*(B^*) = \int_{A^* \cap B^*} \lambda_n^*(x)dx$ for any $B^* \in \mathcal{E}^*$. By (3.53),

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda_n^*)) \leq P(X_1^* \geq u^*) = P(X_1 \geq a_nu^* + b_n) \leq \frac{\varphi(a_nu^* + b_n)}{a_nu^* + b_n},$$

where we used (3.14) for the last inequality. Using (3.8) with $x := u^* \in [-\alpha \log \log n, 0]$ and (3.22), as well as $\alpha \log \log n \leq \log \log n$, we find

$$\begin{aligned} \frac{\varphi(a_nu^* + b_n)}{a_nu^* + b_n} &\leq \frac{e^{-u^*}}{n} \left[1 + \frac{u^* - (\log \log n + \log 4\pi)/2}{2 \log n} \right]^{-1} \\ &\leq \frac{e^{-u^*}}{n} \left[1 - \frac{3 \log \log n + \log 4\pi}{4 \log n} \right]^{-1}. \end{aligned}$$

By using (3.24) and noting that $(6 \log \log n + 2 \log 4\pi)/\log n \leq 5$ for all $n \geq 5$, we obtain the error estimate. (b) By Proposition 2.5.10,

$$\begin{aligned} d_{TV}(\text{PRM}(\lambda_n^*), \text{PRM}(\lambda^*)) &\leq \int_{E^*} |\lambda_n^* - \lambda^*|(dx) \\ &= \int_{u^*}^{\infty} |na_n\varphi(a_nx + b_n) - e^{-x}| dx, \end{aligned}$$

where

$$na_n\varphi(a_nx + b_n) = e^{-x} \cdot \exp \left[-\frac{(2x - \log \log n - \log 4\pi)^2}{16 \log n} \right] \leq e^{-x}.$$

Using $1 - e^{-z} \leq z$ for $z \geq 0$, and $x \in [-\alpha \log \log n, 0]$, we find

$$\begin{aligned} 0 \leq e^{-x} - na_n\varphi(a_nx + b_n) &\leq \frac{(2x - \log \log n - \log 4\pi)^2}{16 \log n} e^{-x} \\ &\leq \frac{(3 \log \log n + \log 4\pi)^2}{16 \log n} e^{-x}. \end{aligned}$$

It follows that

$$d_{TV}(\text{PRM}(\lambda_n^*), \text{PRM}(\lambda^*)) \leq \frac{(3 \log \log n + \log 4\pi)^2}{16 \log n} e^{-u^*}.$$

□

For maxima of Cauchy-distributed random variables, we had to perform an approximation in two steps in Proposition 3.1.9. This is also the case when approximating an MPPE with Cauchy marks by a Poisson process. The reason for this is that the intensity function $[(\pi/n)^2 + x^2]^{-1}$ of the MPPE, which is the intensity function of the Poisson process we first approximate with, varies with the sample size n . Since $[(\pi/n)^2 + x^2]^{-1} \sim x^{-2}$ as $n \rightarrow \infty$, it then makes sense to further approximate the MPPE by a Poisson process with intensity function x^{-2} .

Proposition 3.2.7. *(Standard Cauchy distribution) For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. standard Cauchy random variables, with probability density function $f(y) = 1/\pi(1 + y^2)$ for all $y \in \mathbb{R}$. Define the normalised random variables $X_i^* = \pi n^{-1} X_i$, $i = 1, \dots, n$, taking values in $E^* = \mathbb{R}$. Let $A^* = [u^*, \infty)$ for any choice of $u^* > 0$, and let $\Xi_{A^*}^*$ be defined as in (3.50). Then the mean measure of $\Xi_{A^*}^*$ is given by*

$$\lambda_n^*(B^*) = \int_{A^* \cap B^*} \frac{1}{(\pi/n)^2 + x^2} dx, \quad \text{for any } B^* \in \mathcal{E}^*,$$

and

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda_n^*)) \leq \frac{1}{nu^*}.$$

Moreover, define

$$\lambda^*(B^*) = \int_{A^* \cap B^*} x^{-2} dx, \quad \text{for any } B^* \in \mathcal{E}^*.$$

Then,

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq \frac{1}{nu^*} + \frac{\pi^2}{3n^2 u^{*3}}. \quad (3.56)$$

Proof. We have $a_n = n\pi^{-1}$ and $b_n \equiv 0$. By (3.54) and (3.55), $f^*(x) = n/(\pi^2 + n^2 x^2)$ and $\lambda_n^*(x) = 1/[(\pi/n)^2 + x^2]$ for all $x \in \mathbb{R}$, and $\lambda_n^*(B^*) = \int_{A^* \cap B^*} \lambda_n^*(x) dx$ for any $B^* \in \mathcal{B}(\mathbb{R})$. By (3.53),

$$\begin{aligned} d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda_n^*)) &\leq P(X_1^* \geq u^*) = P\left(X_1 \geq \frac{nu^*}{\pi}\right) \\ &= \bar{F}\left(\frac{nu^*}{\pi}\right) \leq \frac{1}{nu^*}, \end{aligned}$$

where we used (3.32) for the last inequality. Furthermore, note that λ_n^* and λ^* are two finite measures on $E^* = \mathbb{R}$, since

$$\lambda_n^*(E^*) = nP(X_1^* \geq u^*) \leq \frac{1}{u^*} < \infty \quad \text{and} \quad \lambda^*(E^*) = \int_{u^*}^{\infty} x^{-2} dx = \frac{1}{u^*} < \infty.$$

By Proposition 2.5.10, we then have

$$\begin{aligned} d_{TV}(\text{PRM}(\lambda_n^*), \text{PRM}(\lambda^*)) &\leq \int_{E^*} |\lambda_n^* - \lambda^*|(dx) \\ &= \int_{u^*}^{\infty} \left| \frac{1}{(\pi/n)^2 + x^2} - \frac{1}{x^2} \right| dx \\ &= \left(\frac{\pi}{n}\right)^2 \int_{u^*}^{\infty} \frac{1}{(\pi/n)^2 x^2 + x^4} dx. \end{aligned}$$

Since $x \geq u^* > 0$ in the above integral, we obtain the upper bound

$$\left(\frac{\pi}{n}\right)^2 \int_{u^*}^{\infty} x^{-4} dx = \frac{\pi^2}{3n^2 u^{*3}}.$$

□

The expected number of threshold exceedances for the MPPE with Cauchy marks is roughly $1/u_n^*$. The smaller we choose u_n^* (i.e. the closer to 0), the bigger the expected number of threshold exceedances and the smaller the error estimate. Note that for all $u_n^* \leq \pi/\sqrt{3n}$, the second of the two error terms in (3.56) is the bigger one. As an example, choose $u_n^* = 1/\sqrt{n}$. We then expect the MPPE to capture about \sqrt{n} points in $[u_n^*, \infty)$, and the error of the approximation by $\text{PRM}(\int_{A^* \cap \cdot} x^{-2} dx)$ is bounded by

$$\frac{1}{\sqrt{n}} + \frac{\pi^2}{3\sqrt{n}} \leq \frac{4.3}{\sqrt{n}}.$$

3.2.2 Application to MPPE's with geometric marks

In Proposition 3.1.19 we demonstrated that for maxima of geometric random variables, the approximation by a discretised Gumbel distribution living on lattice points k^* gives a smaller error than the approximation by a continuous Gumbel distribution on \mathbb{R} . For the latter approximation to be sharp, we need the condition that the failure probability q_n depends on n in such a way that

$1 - q_n = o(1/\log n)$ for $n \rightarrow \infty$. We encounter a similar behaviour when approximating an MPPE with geometric marks by a Poisson process. Proposition 3.2.8 below gives the error in total variation of the approximation by a Poisson process with mean measure living on the lattice E^* of normalised points k^* . On the other hand, Proposition 3.2.9 determines the error of the approximation by a Poisson process with an easy-to-use continuous mean measure, and uses the d_2 -metric to achieve this.

Proposition 3.2.8. (*Geometric distribution*) *For each integer $n \geq 1$, let X_1, \dots, X_n be i.i.d. geometric random variables with failure probability $q \in (0, 1)$ and $P(X_1 \geq y) = q^{\lceil y \rceil}$, for any $y \geq 0$. Define the normalised random variables $X_i^* = \log(1/q)X_i - \log n$, $i = 1, \dots, n$, taking values in $E^* = \log(1/q)\mathbb{Z}_+ - \log n$. Let $A^* = [u^*, \infty)$ for any choice of $u^* \in [-\log n, \infty)$, and let $\Xi_{A^*}^*$ be defined as in (3.50). Then the mean measure of $\Xi_{A^*}^*$ is given by*

$$\pi^*(B^*) = \sum_{k^* \in A^* \cap E^* \cap B^*} (1 - q)e^{-k^*}, \quad \text{for any } B^* \in \mathcal{B}([-\log n, \infty)), \quad (3.57)$$

and

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\pi^*)) \leq \frac{e^{-u^*}}{n}.$$

Proof. For all $k \in \mathbb{Z}_+$, we use the normalisation $k = (k^* + \log n)/\log(1/q)$, where $k^* \in E^* = \{-\log n, \log(1/q) - \log n, 2\log(1/q) - \log n, \dots\}$. We then have

$$P(X_1 = k) = (1 - q)q^k = (1 - q) \frac{e^{-k^*}}{n} = P(X_1^* = k^*), \quad (3.58)$$

and, for any $B^* \in \mathcal{B}([-\log n, \infty))$,

$$\begin{aligned} \pi^*(B^*) &= nP(X^* \in A^* \cap B^*) \\ &= \sum_{k \in A^* \cap E^* \cap B^*} nP(X_1^* = k^*) \\ &= \sum_{k \in A^* \cap E^* \cap B^*} (1 - q)e^{-k^*}. \end{aligned} \quad (3.59)$$

Using (3.53) and (3.36), we obtain

$$\begin{aligned}
 d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\pi^*)) &\leq P(X_1^* \geq u^*) \\
 &= P\left(X_1 \geq \frac{u^* + \log n}{\log(1/q)}\right) \\
 &= q^{\left\lceil \frac{u^* + \log n}{\log(1/q)} \right\rceil} \leq \frac{e^{-u^*}}{n}.
 \end{aligned}$$

□

The upper error bound that we obtain here is exactly the same as the error bound that we determined in Proposition 3.2.3 for an MPPE with exponential marks, which makes sense as the exponential distribution is the continuous analogue of the geometric distribution. To see this, set $\lambda = \log(1/q)$, and let $Z \sim \text{Exp}(\lambda)$ and $\tilde{Z} \sim \text{Geo}(1 - e^{-\lambda})$. Then $P(Z \geq \lceil z \rceil) = e^{-\lambda \lceil z \rceil} = P(\tilde{Z} \geq z)$, for all $z \geq 0$.

The following proposition now uses the d_2 -metric to approximate the MPPE with geometric marks by a Poisson process with continuous intensity, as the total variation metric is too strong to achieve this. The continuous intensity measure we aim for is the same as that of the MPPE with exponential marks. The result is achieved in two steps: we first use (2.42) to estimate the error in the d_2 -distance of the approximation by a Poisson process with mean measure given by (3.57), and then compare this Poisson process by another one with the desired continuous mean measure, again in the d_2 -distance, by making use of Proposition 2.6.7. We assume here that d_0 is the Euclidean distance on \mathbb{R} bounded by 1, i.e. $d_0(z_1, z_2) = \min(|z_1 - z_2|, 1)$ for any $z_1, z_2 \in \mathbb{R}$, and define the d_1 - and d_2 -distances as in (2.37) and (2.41), respectively, in Section 2.6.

Proposition 3.2.9. (*Geometric distribution*) *For each integer $n \geq 1$, let X_i , X_i^* , $i = 1, \dots, n$, and E^* be defined as in Proposition 3.2.8. Let $A^* = [u^*, \infty)$ for any choice of $u^* \in E^*$, let $\Xi_{A^*}^*$ be defined as in (3.50), with mean measure π^* as in (3.57), and define the continuous measure*

$$\lambda^*(B^*) = \int_{A^* \cap B^*} e^{-x} dx, \quad \text{for any } B^* \in \mathcal{B}([-\log n, \infty)).$$

Then

$$d_2(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq \frac{e^{-u^*}}{n} + 2 \{\log(1/q) \wedge 1\}.$$

Proof. We have

$$d_2(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\lambda}^*)) \leq d_2(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\pi}^*)) + d_2(\text{PRM}(\boldsymbol{\pi}^*), \text{PRM}(\boldsymbol{\lambda}^*)),$$

where, by (2.42) and Proposition 3.2.8,

$$d_2(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\pi}^*)) \leq d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\pi}^*)) \leq \frac{e^{-u^*}}{n}.$$

It thus remains to determine an estimate of $d_2(\text{PRM}(\boldsymbol{\pi}^*), \text{PRM}(\boldsymbol{\lambda}^*))$. Since

$$\boldsymbol{\lambda}^*(A^*) = \int_{u^*}^{\infty} e^{-x} dx = e^{-u^*} = \boldsymbol{\pi}^*(A^*),$$

Proposition 2.6.7 gives

$$\begin{aligned} d_2(\text{PRM}(\boldsymbol{\pi}^*), \text{PRM}(\boldsymbol{\lambda}^*)) &\leq \left(1 - e^{-e^{-u^*}}\right) \left(2 - e^{-e^{-u^*}}\right) d_1(\boldsymbol{\pi}^*, \boldsymbol{\lambda}^*) \\ &\leq 2d_1(\boldsymbol{\pi}^*, \boldsymbol{\lambda}^*). \end{aligned} \quad (3.60)$$

By definition (2.37) of the d_1 -distance,

$$d_1(\boldsymbol{\pi}^*, \boldsymbol{\lambda}^*) = e^{u^*} \sup_{\kappa \in \mathcal{K}} \frac{1}{s_1(\kappa)} \left| \int_{-\log n}^{\infty} \kappa(x) \boldsymbol{\pi}^*(dx) - \int_{-\log n}^{\infty} \kappa(x) \boldsymbol{\lambda}^*(dx) \right|. \quad (3.61)$$

We may write the two integrals in the above expression as a sum of integrals over the “normalised unit intervals” $[k^*, (k+1)^*) = [k^*, k^* + \log(1/q))$, for all $k^* \in E^* \cap [u^*, \infty)$. The modulus then equals

$$\left| \sum_{k^* \geq u^*} \left\{ \int_{k^*}^{k^* + \log(1/q)} \kappa(x) \boldsymbol{\pi}^*(dx) - \int_{k^*}^{k^* + \log(1/q)} \kappa(x) \boldsymbol{\lambda}^*(dx) \right\} \right|. \quad (3.62)$$

Since $\boldsymbol{\pi}^*$ is concentrated on the lattice points $k^* \in E^* \cap [u^*, \infty)$, we have

$$\int_{k^*}^{k^* + \log(1/q)} \kappa(x) \boldsymbol{\pi}^*(dx) = \kappa(k^*) \boldsymbol{\pi}^*({k^*}) = \kappa(k^*) (1-q) e^{-k^*}.$$

Note that we obtain the same result by computing

$$\int_{k^*}^{k^* + \log(1/q)} \kappa(k^*) \boldsymbol{\lambda}^*(dx) = \kappa(k^*) \int_{k^*}^{k^* + \log(1/q)} e^{-x} dx = \kappa(k^*) (1-q) e^{-k^*}.$$

We may thus express (3.62) as follows:

$$\begin{aligned} \left| \sum_{k^* \geq u^*} \int_{k^*}^{k^* + \log(1/q)} \{\kappa(k^*) - \kappa(x)\} \lambda^*(dx) \right| \\ \leq \sum_{k^* \geq u^*} \int_{k^*}^{k^* + \log(1/q)} |\kappa(k^*) - \kappa(x)| \lambda^*(dx). \end{aligned}$$

From Lipschitz continuity of κ , we know that $|\kappa(k^*) - \kappa(x)| \leq s_1(\kappa)d_0(k^*, x)$ for any $x \in [k^*, k^* + \log(1/q))$, where $k^* \in E^* \cap [u^*, \infty)$. The maximum Euclidean distance between k^* and any point in $[k^*, k^* + \log(1/q))$ is of course given by $\log(1/q)$. Since we bound d_0 by 1, we have

$$|\kappa(k^*) - \kappa(x)| \leq s_1(\kappa) \{\log(1/q) \wedge 1\}.$$

For the d_1 -distance in (3.61) we now find, using $\lambda^*([u^*, \infty)) = e^{-u^*}$,

$$d_1(\pi^*, \lambda^*) \leq e^{u^*} \sum_{k^* \geq u^*} \int_{k^*}^{k^* + \log(1/q)} \{\log(1/q) \wedge 1\} \lambda^*(dx) = \log(1/q) \wedge 1,$$

which we plug into (3.60) to obtain an estimate for $d_2(\text{PRM}(\pi^*), \text{PRM}(\lambda^*))$. \square

The approximation of $\mathcal{L}(\Xi_{A^*}^*)$ by $\text{PRM}(\lambda^*)$, whose continuous intensity function e^{-x} corresponds to that of MPPE's with exponential marks, gives rise to an additional error term which depends only on the failure probability of the geometric distribution. With threshold values similar to those that might be used for MPPE's with exponential marks, the error will still become small only if we allow the failure probability $q = q_n$ to tend to 1 as $n \rightarrow \infty$. Since $\log(1/q_n)$ is the length of the normalised unit intervals, this condition causes the lattice structure to melt into the whole real subset $[-\log n, \infty)$ as $n \rightarrow \infty$. Note that Proposition 3.2.9 does not require q_n to vary at a particular rate. The reason for that is that we chose the threshold u_n^* as element of the lattice E^* . If we had not done so, we would have obtained an additional error term of size $\log(1/q_n)e^{-u_n^*}$. In this case, q_n would have needed to vary at a fast enough rate to guarantee a small error despite the factor $e^{-u_n^*}$, which roughly corresponds to the expected number of exceedances and should thus be ≥ 1 . We refer to Section 4.4.5, where we established the error estimate in full detail for MPPE's with bivariate geometric marks.

3.2.3 Remarks on the choice of the point process and its approximation by a Poisson process

Throughout Section 3.2, we have first approximated the law of an MPPE, as defined in (3.45), by a Poisson process with mean measure equal to that of the MPPE. If the mean measure was easy to work with, we were done; else, we approximated further by another process with an easier-to-use mean measure. We will continue to do this for MPPE's with multivariate marks in Chapter 4. The estimate for the first step, the actual "Poisson approximation", comes easily in both chapters. The reason for this is that we use i.i.d. samples X_1, \dots, X_n and i.i.d. indicators $I_{\{X_1 \in A\}}, \dots, I_{\{X_n \in A\}}$. This allows us to apply Proposition 3.2.2, which reduces the problem to the approximation of a binomial by a Poisson distribution. Our main effort, in both chapters, thus lies in determining error bounds on the approximation of a Poisson process by another Poisson process. As the error given by Proposition 3.2.2 is only $P(X \in A)$, the error obtained by further approximating by a different Poisson process is typically the bigger of the two. This might, however, not be the case, if we had a different basic set-up, i.e. if the point process that we consider were different to the MPPE in (3.45). The error arising from approximation by a Poisson process with equal mean measure might then be bigger, and the error from further approximation by a different Poisson process (if not made redundant entirely) might be smaller. For an example of a different basic set-up, assume that we have indicator variables I_i that are dependent, but independent of i.i.d. marks X_i , and let $\Gamma_i^s, \Gamma_i^w, Z_i$ and η_i be defined as in Theorem 2.4.6. We may then apply Theorem 10.H in Barbour et al. (1992) to determine a bound on the approximation, in the total variation distance, of the law of $\Xi = \sum_{i=1}^n I_i \delta_{X_i}$ by $\text{PRM}(\mathbb{E}\Xi)$. A process such as Ξ might appear, for instance, in an insurance context, when considering a claim distribution that is a mixture of typical and large claim sizes. The indices of the occurrences of the large claims may then be dependent, as there may be underlying events leading to these large claims, but the large claim sizes may still be i.i.d.

4 Poisson process approximation for multivariate extremes

The previous chapter gave a first treatment of random configurations of extreme points in space. It dealt with the one-dimensional case, where we considered Poisson process approximation for marked point processes of exceedances (that we called MPPE's) whose marks were univariate. Chapter 4 now studies multivariate extremes. More precisely, instead of i.i.d. random variables as marks, we now consider random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ that are i.i.d. copies of a d -dimensional random vector $\mathbf{X} = (X_1, \dots, X_d)$, where $d \geq 1$. Random point configurations of multivariate extreme points can be modelled, analogously to (3.45) in Chapter 3, by MPPE's of the form

$$\Xi_A = \sum_{i=1}^n I_{\{\mathbf{X}_i \in A\}} \delta_{\mathbf{X}_i}.$$

We suppose that the state space E of the random vectors $\mathbf{X}, \mathbf{X}_1, \dots, \mathbf{X}_n$ is a subset of \mathbb{R}^d and let $\mathcal{E} = \mathcal{B}(E)$. Denote by F the joint distribution of the random vectors and by F_1, \dots, F_d the marginal distribution functions of their components, i.e. for any $\mathbf{y} = (y_1, \dots, y_d) \in E$, let

$$F(\mathbf{y}) = P(\mathbf{X}_i \leq \mathbf{y}) = P(\mathbf{X} \leq \mathbf{y}) = P(X_1 \leq y_1, \dots, X_d \leq y_d),$$

and let

$$F_j(y_j) = P(X_{ij} \leq y_j) = P(X_j \leq y_j),$$

for all $i = 1, \dots, n$ and $j = 1, \dots, d$. Moreover, denote by x_{F_1}, \dots, x_{F_d} the right endpoints of F_1, \dots, F_d , respectively. We fix a set $A \in \mathcal{E}$ such that points (that is, realisations of the random vectors) $\mathbf{x}_i = (x_{i1}, \dots, x_{id})$ lying in it can be considered to be extreme points. In contrast to the univariate case, where it is clear that the set A should be of the form $[u, x_F] \cap E$ (for a certain choice of a threshold u) when studying right-tail extremes, there is more flexibility as to the choice of “extreme region” A in the multivariate case. We might, for

instance, set $A := \{[u_1, x_{F_1}] \times \dots \times [u_d, x_{F_d}]\} \cap E$ which implies that points in A are extreme in *all* components. We might also define A as the complement of $(-\infty, u_1) \times \dots \times (-\infty, u_d)$; then A not only contains jointly extreme points but also points that might have only one extreme component. A similar possibility would be to take a d -dimensional ball of a certain radius $r > 0$ centred in $\mathbf{0}$ and let A be the intersection of $[0, \infty)^d$ with the complement of the ball (or, if looking at all kinds of extreme points, i.e. not only those in the right tail of the marginal distributions, just let A be the complement of the ball). Figure 4.1 illustrates these three particular choices for A in the bivariate case $E = \mathbb{R}_+^2$.

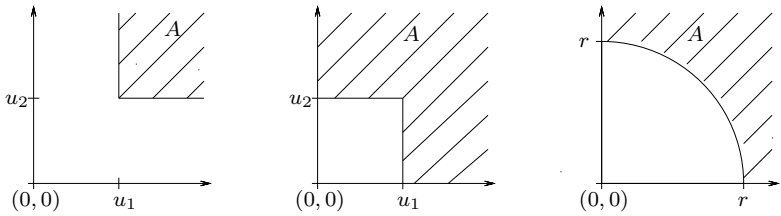


Figure 4.1: We give three examples of choices of the set A in the state space $E = \mathbb{R}_+^2$ (for the case $x_{F_j} = \infty$ for all $j = 1, \dots, d$). (Left) $A = [u_1, \infty) \times [u_2, \infty)$. (Middle) $A = ([0, u_1] \times [0, u_2])^C$. (Right) $A = \{(y_1, y_2) \in E : y_1^2 + y_2^2 \geq r^2\}$.

Analogously to the univariate case (see Proposition 3.2.2), we can apply Theorem 2.5.13 to MPPE's with multivariate marks in order to determine the error in total variation of the approximation of the law of Ξ_A by that of a Poisson process with mean measure $\mathbb{E}\Xi_A$:

Theorem 4.0.1. *For each integer $n \geq 1$, let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be i.i.d. copies of a d -di-mensional random vector \mathbf{X} with state space $E \subseteq \mathbb{R}^d$, where $d \geq 1$. For a fixed set $A \in \mathcal{E}$, let $\Xi_A = \sum_{i=1}^n I_{\{\mathbf{X}_i \in A\}} \delta_{\mathbf{X}_i}$ be the marked point process of points in A and let $W_A = \sum_{i=1}^n I_{\{\mathbf{X}_i \in A\}}$ denote the random number of points in A . Then,*

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \leq P(\mathbf{X} \in A).$$

□

In a first (bivariate) example, we suppose that the components of the random vectors are standard uniformly distributed and independent of each other. More

precisely, we denote the random vectors (or *random pairs*, since $d = 2$) by $\mathbf{U}_1, \dots, \mathbf{U}_n$ and suppose that they are i.i.d. copies of $\mathbf{U} = (U, V)$, where $U, V \sim \text{U}(0, 1)$, and where

$$P(U \leq u, V \leq v) = P(U \leq u)P(V \leq v) = uv. \quad (4.1)$$

The state space of the random vectors is $E = [0, 1]^2$. We fix $A = A_n = [u_n, 1] \times [v_n, 1]$ for some choices of thresholds $u_n, v_n \in [0, 1)$ and introduce the following normalisation:

For $u_n, v_n \in [0, 1)$, there exist $s_n, t_n \in (0, n]$ such that

$$u_n = 1 - \frac{s_n}{n} \text{ and } v_n = 1 - \frac{t_n}{n}.$$

Thus, $(u, v) \in [u_n, 1] \times [v_n, 1] \subseteq [0, 1]^2$ is equivalent to $(s, t) = (n(1-u), n(1-v)) \in (0, s_n] \times (0, t_n] \subseteq (0, n]^2$. Note that this normalisation is equivalent to the slightly different normalisation that we introduced for the univariate case in Propositions 3.1.4 (c), 3.1.7 and 3.2.5. Suppose, for instance, that $s_n = t_n = \log n$. The probability that both components of \mathbf{U} are jointly extreme is

$$P(\mathbf{U} \in A) = P\left(U \geq 1 - \frac{s_n}{n}, V \geq 1 - \frac{t_n}{n}\right) = \frac{s_n t_n}{n^2} = \left(\frac{\log n}{n}\right)^2. \quad (4.2)$$

The probability of the occurrence of joint extremes is thus very small and Poisson approximation, by Theorem 4.0.1, is very sharp. However, the mean of both $\mathcal{L}(W_A)$ and the approximating Poisson distribution is $nP(\mathbf{U} \in A) = s_n t_n / n = \log^2(n)/n$, which is strictly smaller than 1 for all $n \geq 1$, and tends to zero as $n \rightarrow \infty$. For large n , we therefore expect no joint threshold exceedances, so nothing really happens in A for either $\mathcal{L}(W_A)$ or $\text{Poi}(\mathbb{E}W_A)$, and Poisson approximation has to be good. But clearly, the choice $[u_n, 1] \times [v_n, 1]$ for the set A is not the most sensible one for this example. The probability that *one or both* of the components exceeds a threshold is of a higher order than the probability in (4.2):

$$P\left(\left\{U \geq 1 - \frac{s_n}{n}\right\} \cup \left\{V \geq 1 - \frac{t_n}{n}\right\}\right) = \frac{s_n + t_n}{n} - \frac{s_n t_n}{n^2} = \frac{2 \log n}{n} - \left(\frac{\log n}{n}\right)^2,$$

and we expect about $2 \log n$ joint threshold exceedances. A more suitable choice for A in this example is thus $([0, 1 - s_n/n] \times [0, 1 - t_n/n])^C$. Of course, we

can choose different s_n and t_n . The choice of these values depends, on the one hand, on what expected number of exceedances $nP(\mathbf{U} \in A)$ we wish to consider, and, on the other hand, on what size $P(\mathbf{U} \in A)$ of the error we judge to be sufficiently small. The bigger the allowed number of exceedances, the bigger the error will be and vice versa. The approximation will get sharper the farther the set A moves away from the origin $(0,0)$ as there will be less and less points in A . Thus for, say $A = ([0, 1 - 1/n]^2)^C$, we expect only about two threshold exceedances and the error is $2/n$. See Figure 4.2 for an illustration.

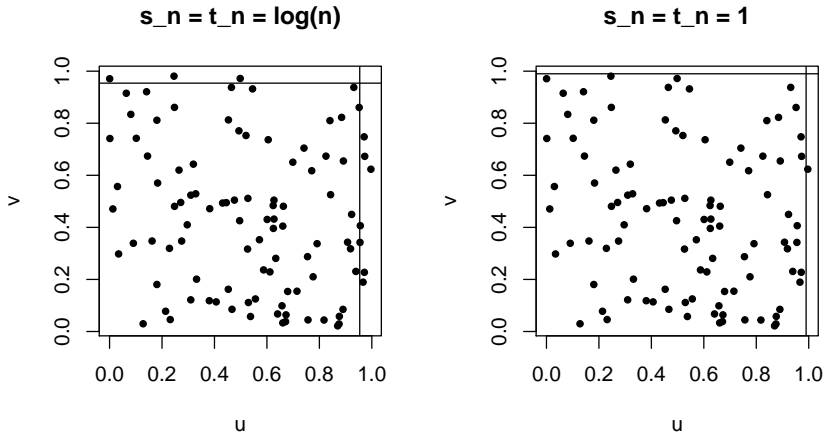


Figure 4.2: We simulate $n = 100$ points from the distribution in (4.1) and compare two sets $A = ([0, 1 - s_n/n] \times [0, 1 - t_n/n])^C$ for different choices of s_n and t_n . Note that we observe no points in the upper-right corner $[1 - s_n/n, 1] \times [1 - t_n/n, 1]$ and that the actual number of threshold exceedances coincides roughly with the expected number of exceedances $\mathbb{E}W_A$. (Left) For the choice $s_n = t_n = \log n$, $\mathbb{E}W_A \approx 2 \log n \approx 9.2$. (Right) For the choice $s_n = t_n = 1$, $\mathbb{E}W_A \approx 2$.

An issue that arises with the use of random vectors as opposed to univariate random variables is thus the question of the choice of A . This issue is connected to the relation between the components of the random vectors: are the components connected in a way to exhibit dependence in the upper tail,

i.e. is there some non-negligible probability of the components being simultaneously extreme? To address this question, we first need to define the joint distribution of the random vectors in more detail, that is, we need to specify the dependence structure between their margins. One way to achieve this is by using copulas. Section 4.1 gives a short introduction to copulas, as well as to bivariate measures of extremal dependence, the so-called *coefficients of tail dependence*. Given that the random vectors have a certain dependence structure, we may then, at least in the case $d = 2$, determine from their coefficient of upper tail dependence whether they are likely to have joint extremes and whether we should define an MPPE with a set A of shape $[u_n, 1) \times [v_n, 1)$. We can turn this idea around. Suppose we are particularly interested in marked point processes with bivariate points that are extreme in both components. Then we may specifically choose copulas that exhibit upper tail dependence for their marks. But why should we actually bother with distinguishing between joint and single-component extremes and not just always use a set A that contains both kinds? The reason for this is that there is an interest, for instance in finance, or in the modelling of extreme events in nature, in providing models for a “perfect storm” scenario, where many things go wrong at the same time.

Another issue that arises with the use of multivariate random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ is that, given certain specified margins, there are infinitely many choices for the joint distribution function and copula. Theorem 4.0.1 fortunately gives a hugely general result for the error that occurs when approximating an process Ξ_A with i.i.d. marks $\mathbf{X}_1, \dots, \mathbf{X}_n$ by a Poisson process with mean measure $\mathbb{E}\Xi_A$, regardless of the common joint distribution of the marks. However, if the joint distribution function has a complicated structure, it might be difficult to understand in what way the error estimate $P(\mathbf{X} \in A)$ varies with the sample size n . To remedy this, Section 4.2 establishes easy bounds on $P(\mathbf{X} \in A)$ for two choices of regions A – one where all components are extreme, and one where there might only be one extreme component – and gives some examples.

Also, the aim should be to approximate a certain choice of an MPPE by a “workable” Poisson process. That is, the Poisson process should have an intensity function that is easy to handle and that preferably does not depend on the sample size n . Whether the Poisson process that we approximate with is useful or not needs to be judged on a case-by-case basis and we thus necessarily need to restrict ourselves to examples. In cases where the intensity function is too difficult to handle, we might try to see if it behaves in a simpler way for $n \rightarrow \infty$ and then approximate by another Poisson process with this simpler intensity, using Proposition 2.5.10. Sections 4.3 and 4.4 each treat

an application of this method. In Section 4.3 we first give a short introduction to the subclass of *Archimedean copulas*. Charpentier and Segers (2009) list examples of bivariate Archimedean copulas according to their asymptotic tail behaviour. Among these, we choose, as examples of possible distribution functions for the marks, those that exhibit asymptotic tail dependence, and show that MPPE's with such marks can be approximated by Poisson processes with practicable intensity functions. In Section 4.4 we study Poisson process approximation for MPPE's with bivariate marks that follow the *Marshall-Olkin geometric distribution*, which is commonly thought of as a natural choice for a bivariate geometric distribution. For this bivariate discrete distribution we of course encounter the same problem as in the univariate case, namely that the total variation distance is too strong if we want to approximate by a Poisson process with a continuous intensity function. We thus use the d_2 -distance as we did before in Section 3.2.2 for the univariate geometric distribution. We contrast our results with those that we obtained in Section 4.2 for the continuous counterpart of this distribution, the *Marshall-Olkin exponential distribution*.

4.1 Copulas and tail dependence

This section gives a very short introduction to copulas and coefficients of tail dependence. A more thorough introduction to copulas along with applications to finance are given in McNeil et al. (2005), a comprehensive treatment can be found in Nelsen (2006), whereas Genest and Nešlehová (2007) treat issues that arise for copulas when using count data. As will be made clear by Theorem 4.1.3 below, copulas come in useful, on the one hand, when trying to understand the dependence structure between the margins of a given distribution function. On the other hand, they are useful for building multivariate models when certain margins are given.

4.1.1 Definition and properties of copulas

Copulas are defined as follows: A d -dimensional *copula* $C : [0, 1]^d \rightarrow [0, 1]$ is a joint distribution function with standard uniform margins. Let $u_1, \dots, u_d \in [0, 1]$. Copulas are characterised by the following three properties:

- (i) $C(u_1, \dots, u_d)$ is increasing in each component u_j , $j = 1, \dots, d$.
- (ii) $C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$, for all $j = 1, \dots, d$ and $u_j \in [0, 1]$.

(iii) For all $(a_1, \dots, a_d), (b_1, \dots, b_d) \in [0, 1]^d$ with $a_j \leq b_j$, we have

$$\sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{i_1+\dots+i_d} C(u_{1i_1}, \dots, u_{di_d}) \geq 0, \quad (4.3)$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j = 1, \dots, d$.

The first property has to be satisfied for any multivariate distribution function, whereas the second property is the requirement of standard uniform margins. The so-called *rectangle inequality* in (4.3) makes sure that $P(a_1 \leq U_1 \leq b_1, \dots, a_d \leq U_d \leq b_d)$ is non-negative for a random vector (U_1, \dots, U_d) with distribution function C .

Example 4.1.1. Let $u, v, u_1, \dots, u_d \in [0, 1]$. We list some examples of well-known copulas; many more examples of copulas can be found in Nelsen (2006).

- (a) *Independence copula:* $\Pi(u_1, \dots, u_d) = \prod_{j=1}^d u_j$. We used this copula in (4.1) for $d = 2$. It is also called the *product copula*.
- (b) *Comonotonicity copula:* $M(u_1, \dots, u_d) = \min_{1 \leq j \leq d} u_j$. This copula is the joint distribution function of a d -dimensional random vector (U, \dots, U) , where U is standard uniformly distributed.
- (c) *Countermonotonicity copula:* $W(u, v) = \max\{u + v - 1, 0\}$. This copula is the joint distribution of $(U, 1 - U)$, where U is standard uniformly distributed.
- (d) *Family of Gumbel(-Hougaard) copulas:* For any $\theta \in [1, \infty)$,

$$C_\theta(u, v) = \exp \left\{ - \left[(-\log u)^\theta + (-\log v)^\theta \right]^{1/\theta} \right\}.$$

- (e) *Family of Clayton copulas:* For any $\theta \in [-1, \infty) \setminus \{0\}$,

$$C_\theta(u, v) = \left[\max(u^{-\theta} + v^{-\theta} - 1, 0) \right]^{-1/\theta}.$$

- (f) *Family of Marshall-Olkin copulas*, also called *family of generalised Cuadras-Augé copulas*: For any $\alpha, \beta \in (0, 1)$,

$$C_{\alpha, \beta}(u, v) = \min(u^{1-\alpha}v, uv^{1-\beta}) = \begin{cases} u^{1-\alpha}v, & u^\alpha \geq v^\beta \\ uv^{1-\beta}, & u^\alpha \leq v^\beta \end{cases}.$$

4.1.2 Fréchet-Hoeffding bounds

The following theorem states that any copula may be bounded by the so-called *Fréchet-Hoeffding bounds* (or *Fréchet bounds*).

Theorem 4.1.2. *For every d -dimensional copula $C(u_1, \dots, u_d)$, we have the bounds*

$$\max \left(\sum_{j=1}^d u_j + 1 - d, 0 \right) \leq C(u_1, \dots, u_d) \leq \min_{1 \leq j \leq d} u_j.$$

Proof. Let C be the joint distribution function of a d -dimensional random vector $\mathbf{U} = (U_1, \dots, U_d)$ with standard uniform margins. For the lower bound, note that, using Boole's inequality $P(\cup_j B_j) \leq \sum_j P(B_j)$ for a countable union of events B_1, B_2, \dots , we obtain

$$\begin{aligned} C(u_1, \dots, u_d) &= P(\cap_{1 \leq j \leq d} \{U_j \leq u_j\}) = 1 - P(\cup_{1 \leq j \leq d} \{U_j > u_j\}) \\ &\geq 1 - \sum_{j=1}^d P(U_j > u_j) = 1 - d + \sum_{j=1}^d u_j, \end{aligned}$$

and remember that a distribution function C is always positive. We obtain the upper bound by noting that, for any $k \in \{1, \dots, d\}$,

$$\bigcap_{1 \leq j \leq d} \{U_j \leq u_j\} \subset \{U_k \leq u_k\}.$$

□

The *Fréchet-Hoeffding upper bound* $\min_{1 \leq j \leq d} u_j$ corresponds to the comonotonicity copula from Example 4.1.1 (b). For $d = 2$, the *Fréchet-Hoeffding lower bound* is precisely the countermonotonicity copula from Example 4.1.1 (c). As shown in Example 5.21 in McNeil et al. (2005), the d -dimensional Fréchet-Hoeffding lower bound is not a copula for $d > 2$, as it does not satisfy the rectangle inequality (4.3). In view of Sklar's Theorem below, the Fréchet bounds may similarly be established for the joint distribution function F of any d -dimensional random vector \mathbf{X} and are expressed in terms of the marginal distribution functions F_1, \dots, F_d of \mathbf{X} :

$$\max \left(\sum_{j=1}^d F_j(y_j) + 1 - d, 0 \right) \leq F(y_1, \dots, y_d) \leq \min_{1 \leq j \leq d} F_j(y_j).$$

4.1.3 Sklar's Theorem

A fundamental result is the following theorem by Sklar (1959), which shows that copulas can be extracted from any joint distribution function. It also shows that a copula, along with some marginal distribution functions, gives all the information that is necessary to define a multivariate joint distribution function.

Theorem 4.1.3. (*Sklar, 1959*)

(i) Let F be a d -dimensional joint distribution function with marginal distribution functions F_1, \dots, F_d . Then there exists a copula $C : [0, 1]^d \rightarrow [0, 1]$ such that for all $y_1, \dots, y_d \in \mathbb{R} = [-\infty, \infty]$,

$$F(y_1, \dots, y_d) = C(F_1(y_1), \dots, F_d(y_d)). \quad (4.4)$$

If the margins are continuous, then the copula C is unique; else C is uniquely determined on $\text{Ran}(F_1) \times \dots \times \text{Ran}(F_d)$, where $\text{Ran}(F_j) = F_j(\mathbb{R})$ denotes the range of F_j , $j = 1, \dots, d$.

(ii) If C is a d -dimensional copula and F_1, \dots, F_d are univariate distribution functions, then F defined by (4.4) is a joint distribution function with margins F_1, \dots, F_d .

Proof. See, for instance, Nelsen (2006) (p. 18-21). □

If the marginal distribution functions F_1, \dots, F_d are continuous, it is easy to obtain (4.4) by using the fact that for a random variable Y with continuous distribution function G , the random variable $G(Y)$ is standard uniformly distributed (see, e.g., Proposition 5.2. in McNeil et al. (2005)):

$$\begin{aligned} F(y_1, \dots, y_d) &= P(X_1 \leq y_1, \dots, X_d \leq y_d) \\ &= P(F_1(X_1) \leq F_1(y_1), \dots, F_d(X_d) \leq F_d(y_d)) \\ &= P(U_1 \leq F_1(y_1), \dots, U_d \leq F_d(y_d)) \\ &= C(F_1(y_1), \dots, F_d(y_d)), \end{aligned}$$

where the random variables U_1, \dots, U_d are standard uniformly distributed and where we let C denote the joint distribution function of the random vector (U_1, \dots, U_d) .

Remark 4.1.4. Note that there is no particular reason or justification behind transforming the marginal distributions to standard uniform distributions, and

thereby, to using copulas. Though copulas provide a way to isolate the dependence structure of a multivariate distribution from its margins and to compare different dependence structures, we could just as well transform the marginal distributions to any other univariate distribution. For instance, multivariate extreme value theory often transforms to standard Fréchet margins, see, e.g. Section 5.4 in Resnick (1987). We choose copulas as one among many possibilities to describe the dependence structure in multivariate distribution functions, mostly out of convenience, as they are well-established in the literature.

Example 4.1.5. Gumbel (1958, 1965) introduced the so-called *Type B bivariate extreme value distribution* as a possible limiting distribution function of the joint distribution function of normalised component-wise maxima; it was later discussed by Kotz et al. (2000) (p. 628) and Nelsen (2006) (p. 28) and is given by

$$F(y_1, y_2) = \exp \left\{ - \left(e^{-\theta y_1} + e^{-\theta y_2} \right)^{1/\theta} \right\},$$

for all $y_1, y_2 \in \mathbb{R}$, where $\theta \in [1, \infty)$. Note that we may rewrite $F(y_1, y_2)$ in the following way in order to see that it satisfies (4.4):

$$\begin{aligned} F(y_1, y_2) &= \exp \left\{ - \left[\left(-\log e^{-e^{-y_1}} \right)^\theta + \left(-\log e^{-e^{-y_2}} \right)^\theta \right]^{1/\theta} \right\} \\ &= C_\theta(F_1(y_1), F_2(y_2)), \end{aligned}$$

where C_θ is the Gumbel copula from Example 4.1.1 (d), and the marginal distributions are standard Gumbel, i.e. $F_1(y) = F_2(y) = \Lambda(y)$, for all $y \in \mathbb{R}$.

4.1.4 Survival copulas

A version of (4.4) also exists for survival functions of multivariate distribution functions. Let \bar{F} be the survival function of a random vector $\mathbf{X} = (X_1, \dots, X_d)$ with state space E , i.e. let

$$\bar{F}(\mathbf{y}) = \bar{F}(y_1, \dots, y_d) = P(X_1 > y_1, \dots, X_d > y_d),$$

for any $\mathbf{y} = (y_1, \dots, y_d) \in E$, and let $F_j, \bar{F}_j = 1 - F_j$ be the marginal and survival functions of X_j , respectively, for each $j = 1, \dots, d$. In the case of continuous margins, $(F_1(X_1), \dots, F_d(X_d))$ is distributed as $(U_1, \dots, U_d) =: \mathbf{U}$,

where the U_j are standard uniform random variables. Denote by C the distribution function of \mathbf{U} and by \hat{C} the *survival copula* of C , that is, let \hat{C} be the joint distribution function of $\mathbf{1} - \mathbf{U}$. Then,

$$\bar{F}(y_1, \dots, y_d) = \hat{C}(\bar{F}_1(y_1), \dots, \bar{F}_d(y_d)), \quad (4.5)$$

since

$$\bar{F}(y_1, \dots, y_d) = P(1 - F_1(X_1) \leq \bar{F}_1(y_1), \dots, 1 - F_d(X_d) \leq \bar{F}_d(y_d)).$$

Note that (4.5) also holds if the margins are discontinuous. Moreover, note that \hat{C} is a copula and not to be mixed up with the survival function of a copula which is not a copula. Denote the survival function of a copula C by \bar{C} . Then

$$\begin{aligned} \bar{C}(u_1, \dots, u_d) &= P(U_1 > u_1, \dots, U_d > u_d) \\ &= P(1 - U_1 \leq 1 - u_1, \dots, 1 - U_d \leq 1 - u_d) \\ &= \hat{C}(1 - u_1, \dots, 1 - u_d). \end{aligned}$$

In the case $d = 2$, we have the following useful relationship between a copula C and its survival copula \hat{C} :

$$\hat{C}(1 - u, 1 - v) = 1 - u - v + C(u, v).$$

Example 4.1.6. Marshall and Olkin (1967a,b, 1985) offer three different derivations of a specific multivariate exponential distribution. One of these is achieved by using a fatal shock model; see also Section 3.1.1 in Nelsen (2006). For simplicity, we suppose that we are in two dimensions. We consider a two-component system, for instance a two engine aircraft. The components fail after they receive a shock (that is always fatal). Let X_1 and X_2 denote the lifetimes of the first and second component, respectively, and let F be the joint distribution function of (X_1, X_2) . We can model the occurrence of shocks to the first, the second, and to both components, up to a time $t \in \mathbb{R}_+$, by three independent Poisson processes $Z_1(t)$, $Z_2(t)$ and $Z_{12}(t)$ with parameters $\nu_1, \nu_2, \nu_{12} > 0$, respectively. For any $y_1, y_2 \in [0, \infty)$, the survival function of (X_1, X_2) is then given by

$$\begin{aligned} \bar{F}(y_1, y_2) &= P(X_1 > y_1, X_2 > y_2) \\ &= P[Z_1(y_1) = 0, Z_2(y_2) = 0, Z_{12}(\max(y_1, y_2)) = 0] \\ &= \exp\{-\nu_1 y_1 - \nu_2 y_2 - \nu_{12} \max(y_1, y_2)\}, \end{aligned} \quad (4.6)$$

and we say that (X_1, X_2) follows the *bivariate Marshall-Olkin exponential distribution*. The survival function of X_1 is given by $\bar{F}_1(y_1) := P(X_1 > y_1) = P[Z_1(y_1) = 0, Z_{12}(y_1) = 0] = \exp\{-(\nu_1 + \nu_{12})y_1\}$, and similarly, the survival function of X_2 is $\bar{F}_2(y_2) = \exp\{-(\nu_2 + \nu_{12})y_2\}$. In order to determine the survival copula \bar{C} , note first that $\max(y_1, y_2) = y_1 + y_2 - \min(y_1, y_2)$ and thus

$$\begin{aligned}\bar{F}(y_1, y_2) &= \exp\{-(\nu_1 + \nu_{12})y_1 - (\nu_2 + \nu_{12})y_2 + \nu_{12} \min(y_1, y_2)\} \\ &= \bar{F}_1(y_1)\bar{F}_2(y_2) \min\{\exp(\nu_{12}y_1), \exp(\nu_{12}y_2)\}.\end{aligned}$$

By setting $u := \bar{F}_1(y_1)$, $v := \bar{F}_2(y_2)$, $\alpha := \nu_{12}/(\nu_1 + \nu_{12}) \in (0, 1)$, and $\beta := \nu_{12}/(\nu_2 + \nu_{12}) \in (0, 1)$, we find that $\exp(\nu_{12}y_1) = u^{-\alpha}$, $\exp(\nu_{12}y_2) = v^{-\beta}$, and therefore

$$\begin{aligned}\bar{F}(y_1, y_2) &= \hat{C}(\bar{F}_1(y_1), \bar{F}_2(y_2)) = \hat{C}(u, v) \\ &= uv \min(u^{-\alpha}, v^{-\beta}) = \min(u^{1-\alpha}v, uv^{1-\beta}).\end{aligned}$$

The survival copula of the bivariate Marshall-Olkin exponential distribution is thus given by the Marshall-Olkin copula that we introduced in Example 4.1.1 (f). The bivariate Marshall-Olkin exponential distribution will be treated in more detail in Section 4.4.

4.1.5 Absolutely continuous and singular components of copulas

A d -dimensional copula need not be absolutely continuous as there might not be a density with respect to Lebesgue measure on \mathbb{R}^d ; see, for instance, Section 2.4 in Nelsen (2006) or Theorem 1.1 in Joe (1997). The copula might thus have a singular component. More precisely (for $d = 2$), each copula may be expressed as follows:

$$C(u, v) = A_C(u, v) + S_C(u, v),$$

where we suppose that A_C is absolutely continuous with respect to two-dimensional Lebesgue measure with density a_C . Then

$$A_C(u, v) = \int_0^v \int_0^u a_C(s, t) \quad \text{and} \quad S_C(u, v) = C(u, v) - A_C(u, v), \quad (4.7)$$

denote the copula's *absolutely continuous* and *singular components*, respectively. If $C \equiv A_C$ and $S_C \equiv 0$ on $[0, 1]^2$, then C is *absolutely continuous* and

$a_C(u, v) = \frac{\partial^2}{\partial v \partial u} C(u, v)$; if $C \equiv S_C$ on $[0, 1]^2$, then C is *singular*. Among the copulas of Example 4.1.1, the independence, Gumbel and Clayton copulas are absolutely continuous, whereas the comonotonicity and countermonotonicity copulas are singular (see also Figure 4.3). The only copula among these to have both an absolutely continuous and a singular component is the Marshall-Olkin copula:

Example 4.1.7. Let $\alpha, \beta \in (0, 1)$ and let $C_{\alpha, \beta}$ belong to the Marshall-Olkin family of copulas as defined in Example 4.1.1 (f). For all $(u, v) \in [0, 1]^2$,

$$\frac{\partial^2}{\partial u \partial v} C_{\alpha, \beta}(u, v) = \begin{cases} (1 - \alpha)u^{-\alpha}, & u^\alpha > v^\beta, \\ (1 - \beta)v^{-\beta}, & u^\alpha < v^\beta, \end{cases} \quad (4.8)$$

and the density $a_C(u, v)$ is given by the right-hand side of (4.8). Integration of a_C as in (4.7) gives the absolutely continuous component of $C_{\alpha, \beta}$:

$$A_C(u, v) = u^{1-\alpha}v - \frac{\alpha\beta}{\alpha + \beta - \alpha\beta} (v^\beta)^{\frac{\alpha+\beta-\alpha\beta}{\alpha\beta}}, \quad \text{for } u^\alpha > v^\beta,$$

whereas $A_C(u, v) = uv^{1-\beta} - \frac{\alpha\beta}{\alpha + \beta - \alpha\beta} (u^\alpha)^{\frac{\alpha+\beta-\alpha\beta}{\alpha\beta}}, \quad \text{for } u^\alpha < v^\beta.$

Thus, for all $(u, v) \in [0, 1]^2$,

$$A_C(u, v) = C_{\alpha, \beta}(u, v) - \frac{\alpha\beta}{\alpha + \beta - \alpha\beta} \left\{ \min(u^\alpha, v^\beta)^{\frac{\alpha+\beta-\alpha\beta}{\alpha\beta}} \right\}, \quad (4.9)$$

and $C_{\alpha, \beta}$ has a singular component concentrated on the curve $u^\alpha = v^\beta$ in $[0, 1]^2$. By (4.7) and (4.9), the singular component is given by

$$S_C(u, v) = \frac{\alpha\beta}{\alpha + \beta - \alpha\beta} \left\{ \min(u^\alpha, v^\beta)^{\frac{\alpha+\beta-\alpha\beta}{\alpha\beta}} \right\} = \int_0^{\min(u^\alpha, v^\beta)} t^{\frac{\alpha+\beta-2\alpha\beta}{\alpha\beta}} dt. \quad (4.10)$$

For two standard uniform random variables U and V whose joint distribution function is given by $C_{\alpha, \beta}$, we have

$$P(U^\alpha = V^\beta) = S_C(1, 1) = \frac{\alpha\beta}{\alpha + \beta - \alpha\beta}.$$

4.1.6 Coefficients of tail dependence

There exist a number of measures that can be used to quantify the dependence between the components of a random pair (X_1, X_2) on $E \subseteq \mathbb{R}^2$. Among these are the linear correlation between X_1 and X_2 , rank correlations like Kendall's tau and Spearman's rho, as well as coefficients of tail dependence; see Section 5.2 in McNeil et al. (2005) or Chapter 5 in Nelsen (2006). Among these, we use the *coefficients of tail dependence* of Joe (1993) which measure the strength of the dependence in the tails of the distribution of a random pair. The coefficient of upper tail dependence is defined as the limiting probability (if it exists) of one of the components exceeding its q -quantile, given that the other component exceeds its q -quantile, for $q \rightarrow 1$. The coefficient of lower tail dependence is defined in a similar way, with both components now being in the lower left quadrant instead of in the upper right quadrant of E . We first define the *generalised inverse* G^\leftarrow of a distribution function G on $B \subseteq \mathbb{R}$ as follows: $G^\leftarrow(w) = \inf\{y \in B : G(y) \geq w\}$, for any $w \in [0, 1]$. Moreover, suppose that X_1 and X_2 are continuous random variables with distribution functions F_1 and F_2 , respectively. We define the *coefficient of upper tail dependence* (if it exists in $[0, 1]$) as follows:

$$\lambda_u := \lambda_u(X_1, X_2) = \lim_{q \uparrow 1} P(X_2 > F_2^\leftarrow(q) \mid X_1 > F_1^\leftarrow(q)).$$

If $\lambda_u = 0$, we say that X_1 and X_2 are *asymptotically independent* in the upper tail; if, however, $\lambda_u \in (0, 1]$, we say that they show asymptotic dependence in the upper tail, or *extremal dependence*. Note that we can exchange $\{X_2 > F_2^\leftarrow(q)\}$ and $\{X_1 > F_1^\leftarrow(q)\}$ in the above definition. Similarly, the *coefficient of lower tail dependence* is defined by

$$\lambda_l := \lambda_l(X_1, X_2) = \lim_{q \downarrow 0} P(X_2 \leq F_2^\leftarrow(q) \mid X_1 \leq F_1^\leftarrow(q)),$$

provided there exists a limit $\lambda_l \in [0, 1]$. By Theorem 5.4.2 in Nelsen (2006), for continuous X_1 and X_2 , the coefficients of tail dependence depend only on the unique copula C of the joint distribution of (X_1, X_2) :

$$\begin{aligned} \lambda_u &:= \lambda_u(C) = \lim_{q \uparrow 1} \frac{\hat{C}(1-q, 1-q)}{1-q} = 2 - \lim_{q \uparrow 1} \frac{1-C(q, q)}{1-q}, \\ \lambda_l &:= \lambda_l(C) = \lim_{q \downarrow 0} \frac{C(q, q)}{q}, \end{aligned}$$

where \hat{C} denotes the survival copula. Furthermore, note that we have

$$\lambda_u(\hat{C}) = \lambda_l(C) \quad \text{and} \quad \lambda_l(\hat{C}) = \lambda_u(C), \quad (4.11)$$

since $\hat{C}(1 - q, 1 - q) = C(1 - q, 1 - q)$. The case of discontinuous margins is examined (for upper tail dependence) in Feidt et al. (2010), where Proposition 4 says that the existence of the coefficient of upper tail dependence is not guaranteed unless the marginal distribution functions satisfy (3.34).

Example 4.1.8. For the copulas in Example 4.1.1 (in the case $d = 2$), we can easily compute the coefficients of upper and lower tail dependence:

Copula	λ_l	λ_u
Independence	0	0
Comonotonicity	1	1
Countermonotonicity	\nexists	0
Gumbel	0	$2 - 2^{1/\theta}$
Clayton ($\theta > 0$)	$2^{-1/\theta}$	0
Marshall-Olkin	0	$\min(\alpha, \beta)$

Figure 4.3 illustrates the tail behaviour of these copulas.

4.2 Poisson process approximation for MPPE's with multivariate marks

Theorem 4.0.1 gives $P(\mathbf{X} \in A)$ as error estimate for the approximation, in the total variation distance, of the law of an MPPE with i.i.d. multivariate marks distributed like \mathbf{X} by that of a Poisson process whose mean measure equals that of the MPPE. We now apply this result to two different choices of the set A that denotes the region in the state space containing extreme points: in Section 4.2.1 we define A such that it contains only points having threshold exceedances in all components, whereas Section 4.2.2 allows A to also contain points that exceed thresholds in only one component. In both cases, we establish easy bounds on the error estimate $P(\mathbf{X} \in A)$ that are valid for any choice of multivariate distribution function for the marks, and for any dimension $d \geq 2$. By way of the example of MPPE's with bivariate Marshall-Olkin exponential marks, Section 4.2.3, together with Example 4.2.9 from Section 4.2.2, highlight the importance of the choice of the scaling constant of the normalisation. Depending on the

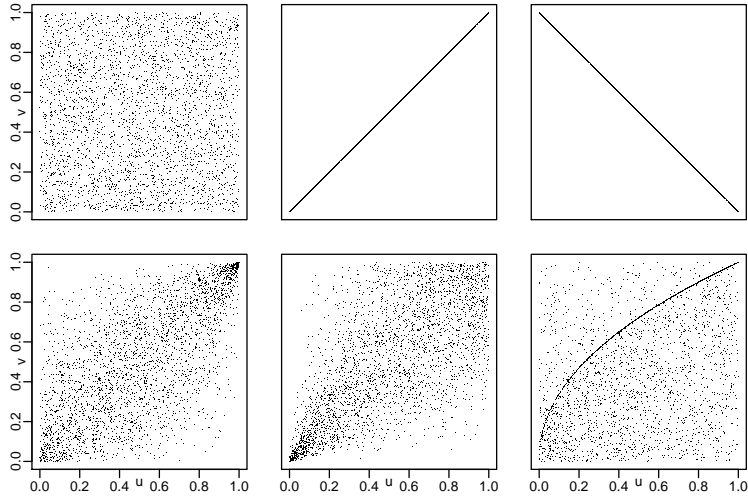


Figure 4.3: We simulate $n = 3000$ points from each of the copulas in Example 4.1.1. (Top left) The independence copula shows no tail dependence. (Top middle) The comonotonicity copula has upper and lower tail dependence concentrated on the line $u = v$. (Top right) The countermonotonicity has no tail dependence; all points are concentrated on the line $v = 1 - u$. (Bottom left) The Gumbel copula with $\theta = 2$ exhibits upper tail dependence. (Bottom middle) The Clayton copula with $\theta = 2$ shows lower tail dependence. (Bottom right) The Marshall-Olkin copula with parameters $\alpha = 0.35$ and $\beta = 0.75$ shows upper tail dependence, concentrated on the curve $u^\alpha = v^\beta$.

scaling, the occurrence of threshold exceedances in both components can have a probability disappearing or non-disappearing with increasing n .

4.2.1 Joint threshold exceedances

We suppose that the possible point configurations taken on by the MPPE's in their d -dimensional state space E only feature points for which every component exceeds a threshold. As discussed in the motivational example at the beginning of Chapter 4, it only makes sense to define a set A and a corresponding MPPE Ξ_A for cases where there is actually some non-negligible probability that the components of the marks are jointly extreme. Theorem 4.2.1 gives an easy error

estimate for any choice of joint distribution function, whereas Proposition 4.2.3 is an application of Theorem 4.2.1 to the case of joint distribution functions with standard uniform margins, i.e. to the case of copulas. Corollary 4.2.2 reformulates Theorem 4.2.1 in terms of normalised random vectors, using the $*$ notation and multivariate analogues of (3.47)-(3.53).

Theorem 4.2.1. *For each integer $n \geq 1$, let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be i.i.d. copies of a d -di-mensional random vector $\mathbf{X} = (X_1, \dots, X_d)$ with state space $E \subseteq \mathbb{R}^d$, joint distribution function F and marginal distribution functions F_j with right endpoints x_{F_j} , respectively, for each $j = 1, \dots, d$, where $d \geq 1$. For a fixed choice of $(u_{1n}, \dots, u_{dn}) \in E$, define*

$$A := A_n := \{(u_{1n}, x_{F_1}] \times \dots \times (u_{dn}, x_{F_d}]\} \cap E.$$

Let $\Xi_A = \sum_{i=1}^n I_{\{\mathbf{X}_i \in A\}} \delta_{\mathbf{X}_i}$ be the marked point process of joint exceedances and let $W_A = \sum_{i=1}^n I_{\{\mathbf{X}_i \in A\}}$ denote the random number of points in A . Then,

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \leq \min_{1 \leq j \leq d} \bar{F}_j(u_{jn}).$$

Proof. By Theorem 4.0.1,

$$\begin{aligned} d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) &\leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \\ &\leq P(\mathbf{X} \in A) = P(X_1 \geq u_{1n}, \dots, X_d \geq u_{dn}) \\ &\leq \min_{1 \leq j \leq d} P(X_j > u_{jn}) = \min_{1 \leq j \leq d} \bar{F}_j(u_{jn}), \end{aligned}$$

where the last inequality follows from

$$\bigcap_{1 \leq j \leq d} \{X_j > u_{jn}\} \subset \{X_k > u_{kn}\}, \quad \text{for any } k \in \{1, \dots, d\}.$$

□

We may thus bound the error of the approximation by the minimum of the marginal probabilities of threshold exceedances (or, in fact, by any of these probabilities, as $\min_{1 \leq j \leq d} \bar{F}_j(u_{jn}) \leq \bar{F}_k(u_{kn})$, for any $k \in \{1, \dots, d\}$). The above theorem might be easier to interpret in applications with random variables subject to a linear transformation; we therefore restate it as follows:

Corollary 4.2.2. *Let $\mathbf{X} = (X_1, \dots, X_d)$ be a d -dimensional random vector with state space $E \subseteq \mathbb{R}^d$, $d \geq 1$. For each integer $n \geq 1$, let $\mathbf{X}_1^*, \dots, \mathbf{X}_n^*$ be i.i.d. copies of the normalised random vector $\mathbf{X}^* = (X_1^*, \dots, X_d^*)$ with state space E^* and joint distribution function F^* , where, for each $j = 1, \dots, d$ and constants $a_{jn}, b_{jn} \in \mathbb{R}$ with $a_{jn} > 0$, the normalised random variable $X_j^* = a_{jn}^{-1}(X_j - b_{jn})$ has distribution function F_j^* and right endpoint $x_{F_j^*}^*$. For a fixed choice of $(u_{1n}^*, \dots, u_{dn}^*) \in E^*$, define*

$$A^* = (u_{1n}^*, x_{F_1^*}^*] \times \dots \times (u_{dn}^*, x_{F_d^*}^*] \cap E^*,$$

let $\Xi_{A^}^* = \sum_{i=1}^n I_{\{\mathbf{X}_i^* \in A^*\}} \delta_{\mathbf{X}_i^*}$ on E^* and $W_{A^*}^* = \sum_{i=1}^n I_{\{\mathbf{X}_i^* \in A^*\}}$. Then,*

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\mathbb{E}\Xi_{A^*}^*)) \leq d_{TV}(\mathcal{L}(W_{A^*}^*), \text{Poi}(\mathbb{E}W_{A^*}^*)) \leq \min_{1 \leq j \leq d} \bar{F}_j^*(u_{jn}^*).$$

□

The following proposition applies Theorem 4.2.1 to copulas.

Proposition 4.2.3. *For each integer $n \geq 1$, let $\mathbf{U}_1, \dots, \mathbf{U}_n$ be i.i.d. copies of a d -dimensional random vector \mathbf{U} with state space $E \subseteq [0, 1]^d$, standard uniform margins U_1, \dots, U_d , and joint distribution function C , where $d \geq 1$. For a fixed choice of $(s_{1n}, \dots, s_{dn}) \in (0, n]^d$, define*

$$A := A_n := \left[1 - \frac{s_{1n}}{n}, 1\right) \times \dots \times \left[1 - \frac{s_{dn}}{n}, 1\right).$$

Let $\Xi_A = \sum_{i=1}^n I_{\{\mathbf{U}_i \in A\}} \delta_{\mathbf{U}_i}$ be the marked point process of joint exceedances and let $W_A = \sum_{i=1}^n I_{\{\mathbf{U}_i \in A\}}$ denote the random number of points in A . Then,

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \leq \min_{1 \leq j \leq d} \frac{s_{jn}}{n}.$$

Proof. By Theorem 4.2.1, we have

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \leq \min_{1 \leq j \leq d} \bar{F}_j(u_{jn}),$$

where $u_{jn} = 1 - s_{jn}/n$ and $\bar{F}_j(u_{jn}) = 1 - u_{jn}$, for all $j = 1, \dots, d$. □

Remark 4.2.4. Proposition 4.2.3 is not a direct application of Corollary 4.2.2. In terms of the notation of Corollary 4.2.2, the norming constants used for Proposition 4.2.3 correspond to $a_{jn} = -(1/n) < 0$ and $b_{jn} = 1$, for all $j = 1, \dots, d$, and thereby do not satisfy the conditions of the corollary.

Example 4.2.5. For $d = 2$ and for any $(u, v) \in E := [0, 1]^2$, consider the family of copulas

$$C_\theta(u, v) = \max \left\{ 1 - [(1-u)^\theta + (1-v)^\theta]^{1/\theta}, 0 \right\}, \quad \text{where } \theta \in [1, \infty).$$

Its coefficient of tail dependence is given by $\lambda_u = 2 - 2^{1/\theta}$ so that the copula C_θ displays upper tail dependence if $\theta \neq 1$. Suppose that $\theta \in (1, \infty)$ and note that, for u, v close to 1, $1 - [(1-u)^\theta + (1-v)^\theta]^{1/\theta}$ will be positive. Let $(U, V), (U_1, V_1), \dots, (U_n, V_n)$ be i.i.d. random pairs with standard uniform margins and joint distribution function C_θ . We choose $(s_n, t_n) \in (0, n)^2$ and define $A = [1 - s_n/n, 1) \times [1 - t_n/n, 1)$, and the MPPE

$$\Xi_A = \sum_{i=1}^n I_{\{U_i \geq 1 - \frac{s_n}{n}, V_i \geq 1 - \frac{t_n}{n}\}} \delta_{(U_i, V_i)}$$

on $\mathcal{B}([0, 1]^2)$. By Proposition 4.2.3,

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \leq \min \left(\frac{s_n}{n}, \frac{t_n}{n} \right),$$

where $\mathbb{E}W_A = \mathbb{E}\Xi_A([0, 1]^2)$ equals

$$\begin{aligned} & nP \left(U \geq 1 - \frac{s_n}{n}, V \geq 1 - \frac{t_n}{n} \right) \\ &= n \left\{ 1 - P \left(U \leq 1 - \frac{s_n}{n} \right) - P \left(V \leq 1 - \frac{t_n}{n} \right) + C_\theta \left(1 - \frac{s_n}{n}, 1 - \frac{t_n}{n} \right) \right\} \\ &= s_n + t_n - (s_n^\theta + t_n^\theta)^{1/\theta}. \end{aligned}$$

We define the intensity function of the approximating Poisson process as follows:

$$\lambda^*(s, t) := \frac{\partial^2}{\partial s \partial t} \left\{ s + t - (s^\theta + t^\theta)^{1/\theta} \right\} = (\theta - 1)(st)^{\theta-1} (s^\theta + t^\theta)^{\frac{1}{\theta}-2},$$

for any $(s, t) \in (0, s_n] \times (0, t_n] \subseteq (0, n)^2$. Define $A^* = (0, s_n] \times (0, t_n]$ and $E^* := [0, n]^2$. The intensity measure of the approximating Poisson process may thus be expressed in terms of the intensity function on $\mathcal{B}(E^*)$, i.e.

$$\lambda^*(B^*) = \int_{A^* \cap B^*} \lambda^*(s, t) ds dt, \quad \text{for any } B^* \in \mathcal{B}(E^*).$$

A Poisson process with intensity measure λ^* is an example of a Poisson process that is easy to use and thus a good choice for approximating the MPPE. The smaller the choices for (one of) the values s_n, t_n , the sharper the approximation will be and the fewer joint exceedances will be expected. For instance, suppose that $s_n = t_n = \log n$. The error estimate for the total variation distance is then $\log(n)/n$ and we expect $(2 - 2^{1/\theta}) \log n$ joint threshold exceedances (again, for $\theta = 1$, we expect no joint threshold exceedances).

4.2.2 Single-component threshold exceedances

In contrast to Section 4.2.1, we here define the set A of extreme points in a way to capture all upper tail extremes of the components. That is, A not only contains d -dimensional points that are extreme in all d components, but also includes points that are extreme in less than d or even only 1 component. The results of this section may thus be used in general for any multivariate distribution regardless of whether it exhibits some kind of joint upper-tail dependence or not. For that, they are also somewhat less precise: the estimate of the error in the total variation distance for approximation by a Poisson process, given by Theorem 4.2.6, is bounded by d times the maximum marginal survival function, whereas the corresponding bound in Section 4.2.1 is only one time the minimum marginal survival function. Corollary 4.2.7 and Proposition 4.2.8 reformulate Theorem 4.2.6 for the cases of normalised random vectors and copulas, respectively.

Theorem 4.2.6. *For each integer $n \geq 1$, let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be i.i.d. copies of a d -dimensional random vector $\mathbf{X} = (X_1, \dots, X_d)$ with state space $E \subseteq \mathbb{R}^d$, joint distribution function F and marginal distribution functions F_j with right endpoints x_{F_j} , respectively, for each $j = 1, \dots, d$, where $d \geq 1$. For a fixed choice of $(u_{1n}, \dots, u_{dn}) \in E$, define*

$$A := A_n := ((-\infty, u_{1n}] \times \dots \times (-\infty, u_{dn}])^C.$$

Let $\Xi_A = \sum_{i=1}^n I_{\{\mathbf{x}_i \in A\}} \delta_{\mathbf{x}_i}$ be the marked point process of joint exceedances and let $W_A = \sum_{i=1}^n I_{\{\mathbf{x}_i \in A\}}$ denote the random number of points in A . Then,

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \leq \sum_{j=1}^d \bar{F}_j(u_{jn}).$$

Proof. By Theorem 4.0.1, and using Boole's inequality, we have that

$$\begin{aligned}
 d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) &\leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \\
 &\leq P(\mathbf{X} \in A) = P(\{X_1 > u_{1n}\} \cup \dots \cup \{X_d > u_{dn}\}) \\
 &\leq \sum_{j=1}^d P(X_j > u_{jn}) = \sum_{j=1}^d \bar{F}(u_{jn}).
 \end{aligned}$$

□

For random vectors whose components are subject to affine transformations, Theorem 4.2.6 reads as follows:

Corollary 4.2.7. *With the notation from Corollary 4.2.2 and $A^* := A_n^* := ((-\infty, u_{1n}^*] \times \dots \times (-\infty, u_{dn}^*])^C$, we obtain*

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\mathbb{E}\Xi_{A^*}^*)) \leq d_{TV}(\mathcal{L}(W_{A^*}^*), \text{Poi}(\mathbb{E}W_{A^*}^*)) \leq \sum_{j=1}^d \bar{F}_j^*(u_{jn}).$$

□

We apply Theorem 4.2.6 to the case where the multivariate marks of the MPPE's are distributed as copulas:

Proposition 4.2.8. *For each integer $n \geq 1$, let $\mathbf{U}_1, \dots, \mathbf{U}_n$ be i.i.d. copies of a d -dimensional random vector \mathbf{U} with state space $E \subseteq [0, 1]^d$, standard uniform margins U_1, \dots, U_d , and joint distribution function C , where $d \geq 1$. For a fixed choice of $\mathbf{s}_n = (s_{1n}, \dots, s_{dn}) \in (0, n]^d$, define*

$$A := A_n := \left(\left[0, 1 - \frac{s_{1n}}{n}\right) \times \dots \times \left[0, 1 - \frac{s_{dn}}{n}\right) \right)^C.$$

Let $\Xi_A = \sum_{i=1}^n I_{\{\mathbf{U}_i \in A\}} \delta_{\mathbf{U}_i}$ be the marked point process of points in A and let $W_A = \sum_{i=1}^n I_{\{\mathbf{U}_i \in A\}}$ denote the random number of points in A . Then,

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \leq \frac{1}{n} \sum_{j=1}^d s_{jn}.$$

Proof. By Theorem 4.0.1, and using Boole's inequality, we have that

$$\begin{aligned}
 d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) &\leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \\
 &\leq P(\mathbf{U} \in A) \\
 &= P\left(\left\{U_1 \geq 1 - \frac{s_{1n}}{n}\right\} \cup \dots \cup \left\{U_d \geq 1 - \frac{s_{dn}}{n}\right\}\right) \\
 &\leq \sum_{j=1}^d P\left(U_j \geq 1 - \frac{s_{jn}}{n}\right) = \sum_{j=1}^d \frac{s_{jn}}{n}.
 \end{aligned}$$

□

Example 4.2.9. Consider again the example from the beginning of Chapter 4, where the common joint distribution function of i.i.d. marks (U, V) , (U_1, V_1) , \dots , (U_n, V_n) is given by the independence copula $C(u, v) = uv$. For any fixed choice of $(s_n, t_n) \in (0, n]^2$, define

$$\begin{aligned}
 A &= \left(\left[0, 1 - \frac{s_n}{n}\right] \times \left[0, 1 - \frac{t_n}{n}\right] \right)^C \\
 &= \left(\left[1 - \frac{s_n}{n}, 1\right] \times [0, 1) \right) \cup \left([0, 1) \times \left[1 - \frac{t_n}{n}, 1\right] \right).
 \end{aligned}$$

With the normalisation $(s, t) = (n(1 - u), n(1 - v))$ for any $(u, v) \in [0, 1)^2$, this set corresponds to $A^\star = ((0, s_n] \times (0, n]) \cup ((0, n] \times (0, t_n])$. By Proposition 4.2.8,

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \leq \frac{s_n + t_n}{n},$$

where

$$\mathbb{E}W_A = \mathbb{E}\Xi_A([0, 1]^2) = s_n + t_n - \frac{s_n t_n}{n} \sim s_n + t_n, \quad \text{as } n \rightarrow \infty.$$

In examples such as this it does not make sense to search for only a bivariate intensity function as in, say, Example 4.2.5, since we essentially have two univariate problems and the bivariate contribution is negligible. More precisely, with $E^\star = [0, n]^2$, write any $B^\star \in \mathcal{B}([0, n]^2)$ as $B_s^\star \times B_t^\star$, where $B_s^\star, B_t^\star \in \mathcal{B}([0, n])$. Then the intensity measure of $\Xi_{A^\star}^\star$ is given by

$$\lambda_n^\star(B^\star) = \int_0^{s_n} I_{B_s^\star} ds + \int_0^{t_n} I_{B_t^\star} dt + \int_0^{s_n} \int_0^{t_n} \left(-\frac{1}{n}\right) I_{B^\star} ds dt,$$

for any $B^* \in \mathcal{B}([0, n]^2)$, and asymptotically behaves like

$$\lambda^*(B^*) := \int_0^{s_n} I_{B_s^*} ds + \int_0^{t_n} I_{B_t^*} dt.$$

By Proposition 2.5.10,

$$d_{TV}(\text{PRM}(\lambda_n^*), \text{PRM}(\lambda^*)) \leq \int_0^{s_n} \int_0^{t_n} \frac{1}{n} ds dt = \frac{s_n t_n}{n}.$$

Example 4.2.10. Let $\nu_1, \nu_2, \nu_{12} > 0$, let $E = [0, \infty)^2$, and let $\mathbf{X} = (X_1, X_2)$ follow the bivariate Marshall-Olkin exponential distribution that we introduced in Example 4.1.6. Since the survival copula \hat{C} of this distribution is given by the Marshall-Olkin copula (see Example 4.1.1 (f)) which has no lower tail dependence, by (4.11), the Marshall-Olkin exponential distribution has no upper tail dependence, as is also illustrated by Figure 4.4. Define

$$A = ([0, u_{1n}) \times [0, u_{2n}))^C = ([u_{1n}, \infty) \times [0, \infty)) \cup ([0, \infty) \times [u_{2n}, \infty))$$

for some choices of thresholds $u_{1n}, u_{2n} \in E$. Let $\mathbf{X}_1^*, \dots, \mathbf{X}_n^*$ be i.i.d. copies of the normalised random variable

$$\mathbf{X}^* = (X_1^*, X_2^*) = ((\nu_1 + \nu_{12})X_1 - \log n, (\nu_2 + \nu_{12})X_2 - \log n) \quad (4.12)$$

with state space $E^* = [-\log n, \infty)^2$ and marginal survival functions

$$\bar{F}_j^*(x_j) = \bar{F}_j\left(\frac{x_j + \log n}{\nu_j + \nu_{12}}\right) = \frac{e^{-x_j}}{n}, \quad \text{for } x_j \geq -\log n \text{ and } j = 1, 2.$$

Then $A^* = ([u_{1n}^*, \infty) \times [-\log n, \infty)) \cup ([-\log n, \infty) \times [u_{2n}^*, \infty))$ and Corollary 4.2.7 gives

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\mathbb{E}\Xi_{A^*}^*)) \leq d_{TV}(\mathcal{L}(W_{A^*}^*), \text{Poi}(\mathbb{E}W_{A^*}^*)) \leq \frac{e^{-u_{1n}^*} + e^{-u_{2n}^*}}{n}.$$

The Marshall-Olkin exponential distribution behaves in different ways above and below the diagonal $y_1 = y_2$, which corresponds to

$$x_2 = \frac{\nu_2 + \nu_{12}}{\nu_1 + \nu_{12}} x_1 + \frac{\nu_2 - \nu_1}{\nu_1 + \nu_{12}} \log n =: \Delta(x_1), \text{ for any } (x_1, x_2) \in E^*.$$

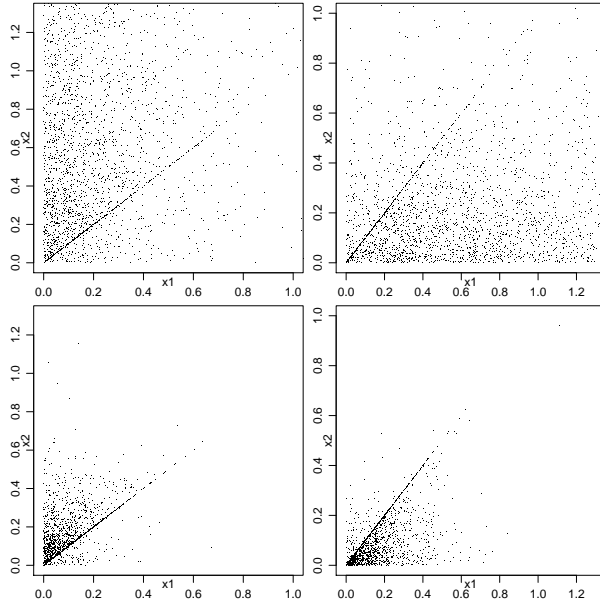


Figure 4.4: We simulate $n = 3000$ points from the Marshall-Olkin exponential distribution. The parameters (ν_1, ν_2, ν_{12}) are given by $(4, 0.5, 0.5)$, $(0.5, 4, 0.5)$, $(4, 0.5, 8)$ and $(0.5, 4, 8)$, respectively (clockwise from top left).

Figure 4.5 shows the three possible shapes for the set A , depending on whether $u_{1n} > u_{2n}$, $u_{1n} = u_{2n}$, or $u_{1n} < u_{2n}$. With $\alpha = \nu_{12}/(\nu_1 + \nu_{12}) \in (0, 1)$, $\beta = \nu_{12}/(\nu_2 + \nu_{12}) \in (0, 1)$, we then have

$$\begin{aligned}
 & \mathbb{E}W_{A^\star}^\star \\
 &= n \{P(X_1^\star \geq u_{1n}^\star) + P(X_2^\star \geq u_{2n}^\star) - P(X_1^\star \geq u_{1n}^\star, X_2^\star \geq u_{2n}^\star)\} \\
 &= n\bar{F}_1^\star(u_{1n}^\star) + n\bar{F}_2^\star(u_{2n}^\star) \\
 &\quad - n \min \left\{ \bar{F}_1^\star(u_{1n}^\star)^{1-\alpha} \cdot \bar{F}_2^\star(u_{2n}^\star), \bar{F}_1^\star(u_{1n}^\star) \cdot \bar{F}_2^\star(u_{2n}^\star)^{1-\beta} \right\} \\
 &= \begin{cases} e^{-u_{1n}^\star} + e^{-u_{2n}^\star} - \left(\frac{1}{n}\right)^{1-\alpha} e^{-(1-\alpha)u_{1n}^\star - u_{2n}^\star} & \text{if } u_{2n}^\star \geq \Delta(u_{1n}^\star) (\Leftrightarrow u_{1n} \leq u_{2n}) \\ e^{-u_{1n}^\star} + e^{-u_{2n}^\star} - \left(\frac{1}{n}\right)^{1-\beta} e^{-u_{1n}^\star - (1-\beta)u_{2n}^\star} & \text{if } u_{2n}^\star \leq \Delta(u_{1n}^\star) (\Leftrightarrow u_{1n} \geq u_{2n}) \end{cases} \\
 &\sim e^{-u_{1n}^\star} + e^{-u_{2n}^\star}, \quad \text{as } n \rightarrow \infty.
 \end{aligned}$$

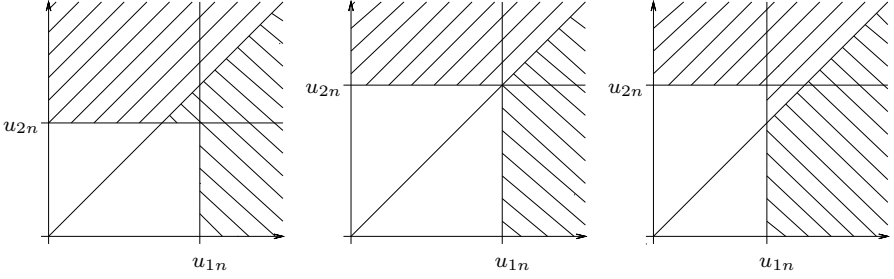


Figure 4.5: Different shapes of the set A from Example 4.2.10 depending on whether $u_{1n} > u_{2n}$ (left), $u_{1n} = u_{2n}$ (middle), or $u_{1n} < u_{2n}$ (right).

The choice of the threshold u_{1n}^* and u_{2n}^* again determines the size of the error estimate and the expected number of threshold exceedances. For instance, for $u_{1n}^* = u_{2n}^* = \log 2 - 0.5 \log(n)$, the error is bounded by $n^{-1/2}$ and the MPPE captures roughly the \sqrt{n} points farthest away from the origin $(-\log n, -\log n)$ of the normalised state space.

Remark 4.2.11. Theorem 4.2.6 is easily adapted to the case where A allows for a selection of the components to be extreme but specifically forbids the remaining components to be so. Let $J \subseteq \{1, \dots, d\}$ and set

$$A := A_n := \left(\prod_{j \in J} (-\infty, u_{jn}] \right)^C \times \left(\prod_{k \in \{1, \dots, d\} \setminus J} (-\infty, u_{kn}] \right).$$

Then, $d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A))$ is bounded by

$$\begin{aligned} P(\mathbf{X} \in A) &= P \left(\bigcup_{j \in J} \{X_j > u_{jn}\}, \bigcap_{k \in \{1, \dots, d\} \setminus J} \{X_k \leq u_{kn}\} \right) \\ &\leq \left(\sum_{j \in J} \bar{F}(u_{jn}) \right) \cdot \left(1 - \min_{k \in \{1, \dots, d\} \setminus J} \bar{F}(u_{kn}) \right) \\ &\leq \sum_{j \in J} \bar{F}(u_{jn}) \leq |J| \max_{j \in J} \bar{F}(u_{jn}). \end{aligned}$$

4.2.3 An example of hidden dependence for joint threshold exceedances

We consider again the bivariate Marshall-Olkin exponential distribution from Examples 4.1.6 and 4.2.10. With the normalisation used in (4.12), the probability of the occurrence of joint extremes is negligible compared against the probability of extremes in the margins. With a different scaling however, we may zoom in on the region of joint extremes and model the behaviour of an MPPE with Marshall-Olkin exponential marks in that region by a Poisson process with an intensity that will need to be determined.

We use the following normalisation of the Marshall-Olkin exponentially distributed random pair $\mathbf{X} = (X_1, X_2)$:

$$\mathbf{X}^* = (X_1^*, X_2^*) = (\nu X_1 - \log n, \nu X_2 - \log n),$$

where $\nu := \nu_1 + \nu_2 + \nu_{12}$. Let $\mathbf{X}_1^*, \dots, \mathbf{X}_n^*$ be i.i.d. copies of \mathbf{X}^* . Their state space is given by $E^* = [-\log n, \infty)^2$, and the marginal survival functions by

$$\bar{F}_j^*(x_j) = \bar{F}_j\left(\frac{x_j + \log n}{\nu}\right) = \left(\frac{e^{-x_j}}{n}\right)^{\frac{\nu_j + \nu_{12}}{\nu}}, \quad \text{for } x_j \geq -\log n \text{ and } j = 1, 2.$$

Let $u_{1n}^*, u_{2n}^* \in E^*$ and define $A^* = A_n^* = (u_{1n}^*, \infty) \times (u_{2n}^*, \infty)$. The expected number of points in A^* is given by

$$\begin{aligned} \mathbb{E}W_{A^*}^* &= n\bar{F}\left(\frac{u_{1n}^* + \log n}{\nu}, \frac{u_{2n}^* + \log n}{\nu}\right) \\ &= \begin{cases} \exp\left(-\frac{\nu_1}{\nu} u_{1n}^* - \frac{\nu_2 + \nu_{12}}{\nu} u_{2n}^*\right), & \text{if } u_{1n}^* < u_{2n}^*, \\ \exp(-u_{1n}^*), & \text{if } u_{1n}^* = u_{2n}^*, \\ \exp\left(-\frac{\nu_1 + \nu_{12}}{\nu} u_{1n}^* - \frac{\nu_2}{\nu} u_{2n}^*\right), & \text{if } u_{1n}^* > u_{2n}^*. \end{cases} \end{aligned} \quad (4.13)$$

By Theorem 4.0.1,

$$\begin{aligned} d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\mathbb{E}\Xi_{A^*}^*)) &\leq d_{TV}(\mathcal{L}(W_{A^*}^*), \text{Poi}(\mathbb{E}W_{A^*}^*)) \\ &\leq P(X_1^* \geq u_{1n}^*, X_2^* \geq u_{2n}^*) = \frac{\mathbb{E}W_{A^*}^*}{n}, \end{aligned} \quad (4.14)$$

with $\mathbb{E}W_{A^*}^*$ from (4.13). We could of course also use Corollary 4.2.2, but it gives a worse error estimate:

$$\begin{aligned} d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\mathbb{E}\Xi_{A^*}^*)) &\leq d_{TV}(\mathcal{L}(W_{A^*}^*), \text{Poi}(\mathbb{E}W_{A^*}^*)) \\ &\leq \min \left\{ \left(\frac{e^{-u_{1n}^*}}{n}\right)^{\frac{\nu_1 + \nu_{12}}{\nu}}, \left(\frac{e^{-u_{2n}^*}}{n}\right)^{\frac{\nu_2 + \nu_{12}}{\nu}} \right\}. \end{aligned}$$

Example 4.2.12. Suppose that $\nu_1 \geq \nu_2$ and that

$$u_{1n}^* = -\frac{\nu}{\nu_1 + \nu_{12}} \log \log n \quad \text{and} \quad u_{2n}^* = -\frac{\nu}{\nu_2 + \nu_{12}} \log \log n.$$

Then $u_{1n}^* \geq u_{2n}^*$ for all $n > 2$, and we expect $(\log n)^{1+\frac{\nu_2}{\nu_2+\nu_{12}}}$ points in A^* . The error estimate in (4.14) is given by

$$\frac{(\log n)^{1+\frac{\nu_2}{\nu_2+\nu_{12}}}}{n}.$$

Example 4.2.13. Suppose that $u_{1n}^* = u_{2n}^* = -\log \log n$. Then we expect $\log n$ points in A^* and the error estimate in (4.14) is $\log(n)/n$.

The Marshall-Olkin exponential distribution has both an absolutely continuous and a singular part (see Theorem 3.1 in Marshall and Olkin (1967b) or deduce it from the underlying Marshall-Olkin survival copula and Example 4.1.7). We denote the intensity functions of the approximating Poisson process in the original and in the normalised state spaces E and E^* by $\lambda(y_1, y_2)$ and $\lambda^*(x_1, x_2)$, respectively, for the absolutely continuous part (i.e. for $y_1 \neq y_2$ and $x_1 \neq x_2$), and by $\acute{\lambda}(y)$ and $\acute{\lambda}^*(x)$, respectively, for the singular part (i.e. for $y_1 = y_2 = y$ and $x_1 = x_2 = x$). The corresponding intensity measures are

$$\lambda_A(B) = \mathbb{E}\Xi_A(B) = \int_{A \cap B} \lambda(y_1, y_2) dy_1 dy_2 + \int_{A \cap B \cap \{(y_1, y_2): y_1 = y_2\}} \acute{\lambda}(y) dy,$$

$$\begin{aligned} \lambda_{A^*}^*(B^*) = \mathbb{E}\Xi_{A^*}^*(B^*) &= \int_{A^* \cap B^*} \lambda^*(x_1, x_2) dx_1 dx_2 \\ &\quad + \int_{A^* \cap B^* \cap \{(x_1, x_2): x_1 = x_2\}} \acute{\lambda}^*(x) dx, \end{aligned}$$

for any $B \in \mathcal{B}(E)$ and for any $B^* \in \mathcal{B}(E^*)$, respectively. We can use (4.8) from Example 4.1.7 with the transformations $u := \bar{F}_1(y_1)$, $v := \bar{F}_2(y_2)$ and $\alpha = \nu_{12}/(\nu_1 + \nu_{12})$, $\beta = \nu_{12}/(\nu_2 + \nu_{12})$ in order to determine the bivariate density function $f(y_1, y_2)$ of the Marshall-Olkin exponential distribution, and thereby also $\lambda(y_1, y_2) = n f(y_1, y_2)$:

$$\lambda(y_1, y_2) = \begin{cases} n\nu_1(\nu_2 + \nu_{12}) \exp\{-\nu_1 y_1 - (\nu_2 + \nu_{12}) y_2\}, & \text{if } y_1 < y_2, \\ n\nu_2(\nu_1 + \nu_{12}) \exp\{-(\nu_1 + \nu_{12}) y_1 - \nu_2 y_2\}, & \text{if } y_1 > y_2. \end{cases}$$

Substitution by $t := \exp(-\nu_{12}w)$ in (4.10) gives $S_C(y, y) = \int_0^y \nu_{12} \exp\{-\nu w\} dw$ and the intensity function on the diagonal is thus given by

$$\dot{\lambda}(y) = n\nu_{12} \exp(-\nu y), \quad \text{for } y_1 = y_2 = y.$$

With the affine transformations

$$\begin{aligned} \phi : E^* &\rightarrow E, \quad (x_1, x_2) \mapsto \phi(x_1, x_2) = \left(\frac{x_1 + \log n}{\nu}, \frac{x_2 + \log n}{\nu} \right), \\ \tau : [-\log n, \infty) &\rightarrow [0, \infty), \quad x \mapsto \tau(x) = \frac{x + \log n}{\nu}, \end{aligned}$$

we obtain $\lambda^*(x_1, x_2) = \nu^{-2} \lambda(\phi(x_1, x_2))$, for $x_1 \neq x_2$, and $\dot{\lambda}^*(x) = \nu^{-1} \dot{\lambda}(\tau(x))$, for $x_1 = x_2 = x$. Hence, for any $(x_1, x_2) \in E^*$,

$$\begin{aligned} \lambda^*(x_1, x_2) &= \begin{cases} \frac{\nu_1(\nu_2 + \nu_{12})}{\nu^2} \exp\left(-\frac{\nu_1}{\nu} x_1 - \frac{\nu_2 + \nu_{12}}{\nu} x_2\right), & \text{if } x_1 < x_2, \\ \frac{\nu_2(\nu_1 + \nu_{12})}{\nu^2} \exp\left(-\frac{\nu_1 + \nu_{12}}{\nu} x_1 - \frac{\nu_2}{\nu} x_2\right), & \text{if } x_1 > x_2, \end{cases} \\ \dot{\lambda}^*(x) &= \frac{\nu_{12}}{\nu} \exp(-x), \quad \text{if } x_1 = x_2 = x. \end{aligned}$$

4.3 Archimedean copulas with upper tail dependence

A well known class of copulas are the so-called *Archimedean copulas*, discussed in detail in, e.g. Chapter 4 in Nelsen (2006). A copula C is called Archimedean copula with *generator* ϕ , if it can be expressed in the following way:

$$C(u_1, \dots, u_d) = \phi^{[-1]}(\phi(u_1) + \dots + \phi(u_d)),$$

where the function $\phi : [0, 1] \rightarrow [0, \infty]$ is continuous, strictly decreasing, convex, and satisfies $\phi(1) = 0$. The *pseudo-inverse* $\phi^{[-1]} : [0, \infty] \rightarrow [0, 1]$ of ϕ is defined as follows:

$$\phi^{[-1]}(x) := \begin{cases} \phi^{-1}(x), & 0 \leq x \leq \phi(0), \\ 0, & \phi(0) \leq x \leq \infty. \end{cases}$$

The copula is called *strict* if $\lim_{r \downarrow 0} \phi(r) = \infty$, and in this case $\phi^{[-1]} = \phi^{-1}$.

Now let $\mathbf{U}_i = (U_i, V_i)$, $i = 1, \dots, n$ be i.i.d. copies of a random pair (U, V) whose joint distribution function is given by a bivariate Archimedean copula,

and let A denote a region in $E = [0, 1)^2$ of joint upper tail extremes. More precisely, we define, as in Proposition 4.2.3,

$$A = A_n = \left[1 - \frac{s_n}{n}, 1\right) \times \left[1 - \frac{t_n}{n}, 1\right), \quad \text{for any choice of } (s_n, t_n) \in (0, n]^2. \quad (4.15)$$

Moreover, define the number W_A of (U_i, V_i) 's in A , as well as the MPPE Ξ_A on E with marks (U_i, V_i) by

$$W_A = \sum_{i=1}^n I_{\{U_i \geq 1 - \frac{s_n}{n}, V_i \geq 1 - \frac{t_n}{n}\}} \quad \text{and} \quad \Xi_A = \sum_{i=1}^n I_{\{U_i \geq 1 - \frac{s_n}{n}, V_i \geq 1 - \frac{t_n}{n}\}} \delta_{(U_i, V_i)}, \quad (4.16)$$

respectively. It follows from Proposition 4.2.3 that

$$d_{TV}(\mathcal{L}(\Xi_A), \text{PRM}(\mathbb{E}\Xi_A)) \leq d_{TV}(\mathcal{L}(W_A), \text{Poi}(\mathbb{E}W_A)) \leq \min\left(\frac{s_n}{n}, \frac{t_n}{n}\right). \quad (4.17)$$

As usual, we can regulate the size of this error estimate by the size of the values s_n and t_n . The smaller $\min(s_n, t_n)$ with respect to the sample size n , the smaller the error of the approximation by the Poisson process. However, it is not necessarily clear how to interpret $\mathbb{E}W_A$ or $\mathbb{E}\Xi_A$, as the structure of the copula can be complicated. We will illustrate this by way of an example below. But first we define, for any $r \in [0, 1]$ and for any $x \in [0, \infty]$,

$$\bar{\phi}(r) := \phi(1 - r) \quad \text{and} \quad \bar{\phi}^{[-1]}(x) := 1 - \phi^{[-1]}(x). \quad (4.18)$$

The expected number of exceedances may then in general be expressed as follows:

$$\begin{aligned} \mathbb{E}W_A &= nP\left(U \geq 1 - \frac{s_n}{n}, V \geq 1 - \frac{t_n}{n}\right) \\ &= s_n + t_n - n\bar{\phi}^{[-1]}\left(\bar{\phi}\left(\frac{s_n}{n}\right) + \bar{\phi}\left(\frac{t_n}{n}\right)\right). \end{aligned}$$

With the normalisation $(s_n, t_n) = (n(1 - u), n(1 - v))$, where $(u, v) \in E$, we obtain

$$A^* = (0, s_n] \times (0, t_n] \subseteq (0, n]^2. \quad (4.19)$$

On any set $B^* \in \mathcal{B}((0, n]^2)$, the intensity measure of the MPPE is then given by

$$\mathbb{E}\Xi_{A^*}^*(B^*) = \int_{A^* \cap B^*} e_n^*(s, t) ds dt,$$

where $e_n^*(s, t)$ is the intensity function given by

$$\begin{aligned} e_n^*(s, t) &= \frac{\partial^2}{\partial t \partial s} nP \left(U \geq 1 - \frac{s}{n}, V \geq 1 - \frac{t}{n} \right) \\ &= \frac{\partial^2}{\partial t \partial s} (-n) \bar{\phi}^{[-1]} \left(\bar{\phi} \left(\frac{s}{n} \right) + \bar{\phi} \left(\frac{t}{n} \right) \right), \end{aligned} \quad (4.20)$$

for any $(s, t) \in (0, s_n] \times (0, t_n] \subseteq (0, n]^2$. We next consider $\mathbb{E}W_A$ and $e_n^*(s, t)$ for a specific example of an MPPE with marks distributed according to an Archimedean copula.

Example 4.3.1. The Gumbel copula from Example 4.1.1 (d) is an Archimedean copula with generator $\phi(r) = (-\log r)^\theta$ and $\theta \geq 1$. As seen in Example 4.1.8, it has upper tail dependence. We assume the notation and setting introduced above. The expected number of joint exceedances is then given by

$$\mathbb{E}W_A = -n + s_n + t_n + n \exp \left\{ - \left[\left(-\ln \left(1 - \frac{s_n}{n} \right) \right)^\theta + \left(-\ln \left(1 - \frac{t_n}{n} \right) \right)^\theta \right]^{\frac{1}{\theta}} \right\}.$$

Moreover, by computing the double derivative in (4.20), we obtain the following intensity function:

$$\begin{aligned} e_n^*(s, t) &= \frac{1}{n} \cdot \frac{(\log(1 - \frac{s}{n}) \log(1 - \frac{t}{n}))^{\theta-1}}{(1 - \frac{s}{n})(1 - \frac{t}{n})} \cdot e^{-[(-\log(1 - \frac{s}{n}))^\theta + (-\log(1 - \frac{t}{n}))^\theta]^{1/\theta}} \\ &\quad \cdot \left[\left(-\log \left(1 - \frac{s}{n} \right) \right)^\theta + \left(-\log \left(1 - \frac{t}{n} \right) \right)^\theta \right]^{\frac{1}{\theta}-2} \\ &\quad \cdot \left\{ \theta - 1 + \left[\left(-\log \left(1 - \frac{s}{n} \right) \right)^\theta + \left(-\log \left(1 - \frac{t}{n} \right) \right)^\theta \right]^{\frac{1}{\theta}} \right\}. \end{aligned}$$

Clearly, it is not evident how many joint threshold exceedances to expect for specific choices of s_n and t_n . Moreover, $e_n^*(s, t)$ is impracticable as intensity function of the approximating Poisson process: on the one hand its structure is too complicated to work with, and, on the other hand, it depends on the sample size n . However, by subsequently using $-\log(1 - w) \sim w$ and $e^{-z} \sim 1 - z$ for $w, z \rightarrow 0$, we find that

$$\mathbb{E}W_A \sim -n + s_n + t_n + n \exp \left\{ - \frac{(s_n^\theta + t_n^\theta)^{1/\theta}}{n} \right\} \sim s_n + t_n - (s_n^\theta + t_n^\theta)^{1/\theta}, \quad (4.21)$$

as $n \rightarrow \infty$. A “nicer” approximative intensity function can thus be defined as follows

$$\lambda^*(s, t) := \frac{\partial^2}{\partial s \partial t} \left\{ s + t - (s^\theta + t^\theta)^{1/\theta} \right\} = (\theta - 1)(st)^{\theta-1} (s^\theta + t^\theta)^{\frac{1}{\theta}-2}, \quad (4.22)$$

for any $(s, t) \in (0, s_n] \times (0, t_n] \subseteq (0, n]^2$. Instead of approximating the law of Ξ_A by a Poisson process with intensity function $e_n^*(s, t)$, the aim would rather be to approximate it by a Poisson process with the simpler intensity function $\lambda^*(s, t)$ and use Proposition 2.5.10 to estimate $d_{TV}(\text{PRM}(\mathbb{E}\Xi_{A^*}^*), \text{PRM}(\lambda^*))$, where

$$\lambda^*(B^*) = \lambda_{A^*}^*(B^*) = \int_{A^* \cap B^*} \lambda^*(s, t) ds dt, \quad \text{for any } B^* \in \mathcal{B}((0, n]^2). \quad (4.23)$$

There is an entire subclass of Archimedean copulas showing the asymptotic behaviour (4.21) for $\mathbb{E}W_A$. This is demonstrated in Section 4.3.1, where an asymptotic result due to Charpentier and Segers (2009) provides a way to determine which Archimedean copulas exhibit upper tail dependence similar to the Gumbel copula. For any such copula, Section 4.3.2 determines a bound on the total variation distance between $\text{PRM}(\mathbb{E}\Xi_{A^*}^*)$ and $\text{PRM}(\lambda^*)$. In Section 4.3.3 we apply our results to a list of examples of Archimedean copulas with upper tail dependence. Section 4.3.4 gives a discussion on ways to determine bounds if $d \geq 3$.

4.3.1 An asymptotic result

Charpentier and Segers (2009) showed that there is a subclass of Archimedean copulas all displaying the asymptotic behaviour that we found in (4.21). They noted that the upper tail behaviour of Archimedean copulas can be determined by computing some characteristics of their generator ϕ . The following theorem, formulated more generally in Charpentier and Segers (2009) (see Theorem 4.1), gives two different possibilities for the asymptotic behaviour, as $n \rightarrow \infty$, of

$$\mathbb{E}W_A = nP \left(U \geq 1 - \frac{s_n}{n}, V \geq 1 - \frac{t_n}{n} \right),$$

depending on the value of $\tilde{\theta}$ that we define in (4.24) below.

Theorem 4.3.2. (Charpentier and Segers, 2009) Let (U, V) be a random pair with standard uniform margins and joint distribution function C , which is a bivariate Archimedean copula with generator ϕ . If the limit

$$\tilde{\theta} := -\lim_{r \downarrow 0} \frac{r\phi'(u)|_{u=1-r}}{\phi(1-r)}, \quad (4.24)$$

exists in $[1, \infty)$, then, for every $(s, t) \in (0, \infty)^2$,

$$\begin{aligned} \lim_{n \rightarrow \infty} nP \left(U \geq 1 - \frac{s}{n}, V \geq 1 - \frac{t}{n} \right) \\ = \begin{cases} 0, & \text{if } \tilde{\theta} = 1, \\ s + t - \left(s^{\tilde{\theta}} + t^{\tilde{\theta}} \right)^{1/\tilde{\theta}}, & \text{if } 1 < \tilde{\theta} < \infty. \end{cases} \end{aligned} \quad (4.25)$$

Remark 4.3.3. Charpentier and Segers (2009) showed that (4.24) is equivalent to regular variation of the function $\bar{\phi}$ at 0 with index $\tilde{\theta}$, i.e. to

$$\lim_{r \downarrow 0} \frac{\bar{\phi}(rx)}{\bar{\phi}(r)} = x^{\tilde{\theta}}, \quad \text{for any } x \in (0, \infty).$$

This fact is used in the proof of Theorem 4.3.2 below.

Proof. We have

$$\begin{aligned} nP \left(U \geq 1 - \frac{s}{n}, V \geq 1 - \frac{t}{n} \right) \\ = n \left[P \left(U \geq 1 - \frac{s}{n} \right) + P \left(V \geq 1 - \frac{t}{n} \right) \right. \\ \quad \left. - P \left(\left\{ U \geq 1 - \frac{s}{n} \right\} \cup \left\{ V \geq 1 - \frac{t}{n} \right\} \right) \right] \\ = s + t - nP \left(\left\{ U \geq 1 - \frac{s}{n} \right\} \cup \left\{ V \geq 1 - \frac{t}{n} \right\} \right), \end{aligned}$$

where

$$\begin{aligned} nP \left(\left\{ U \geq 1 - \frac{s}{n} \right\} \cup \left\{ V \geq 1 - \frac{t}{n} \right\} \right) &= n\bar{\phi}^{[-1]} \left(\bar{\phi} \left(\frac{s}{n} \right) + \bar{\phi} \left(\frac{t}{n} \right) \right) \\ &= \frac{1}{\bar{\phi}^{[-1]}(\bar{\phi}(\frac{1}{n}))} \cdot \bar{\phi}^{[-1]} \left(\bar{\phi} \left(\frac{1}{n} \right) \cdot \left[\frac{\bar{\phi}(\frac{s}{n})}{\bar{\phi}(\frac{1}{n})} + \frac{\bar{\phi}(\frac{t}{n})}{\bar{\phi}(\frac{1}{n})} \right] \right). \end{aligned} \quad (4.26)$$

By Remark 4.3.3, $\bar{\phi}$ is regularly varying at 0 with index $\tilde{\theta}$. It follows that the function $x \mapsto 1/\bar{\phi}(1/x)$ is regularly varying at infinity with index $\tilde{\theta}$. By Theorem 1.5.12 in Bingham et al. (1987), its inverse function, i.e. $y \mapsto 1/\bar{\phi}^{[-1]}(1/y)$ is regularly varying at infinity with index $1/\tilde{\theta}$, and the function $\bar{\phi}^{[-1]}$ is regularly varying at 0 with index $1/\tilde{\theta}$. By the uniform convergence theorem (see Theorem 1.5.2 in Bingham et al. (1987)), (4.26) thus converges to

$$(s^{\tilde{\theta}} + t^{\tilde{\theta}})^{1/\tilde{\theta}}, \quad \text{as } n \rightarrow \infty.$$

□

If $\tilde{\theta} = 1$, the copula displays asymptotic independence in the upper tail. For $1 < \tilde{\theta} < \infty$, it shows upper tail dependence and we recognise the limiting structure of $\mathbb{E}W_A$ in (4.21) that we obtained for the Gumbel example. Indeed, for the Gumbel copula with generator $\phi(r) = (-\log r)^\theta$, we have

$$\begin{aligned} \phi'(u)|_{u=1-r} &= \frac{(-\theta)[-\log(1-r)]^{\theta-1}}{1-r}, \\ \text{and } \tilde{\theta} &= \lim_{r \downarrow 0} \frac{\theta[-\log(1-r)]^{\theta-1}}{(1-r)[-\log(1-r)]^\theta} = \lim_{r \downarrow 0} \frac{\theta r}{(1-r)[-\log(1-r)]} = \theta \in [1, \infty). \end{aligned}$$

For $\theta = 1$, the Gumbel copula reduces to the independence copula, which is obviously asymptotically independent (see Example 4.1.8 for the coefficients of upper tail dependence of the independence and Gumbel copulas). In the following, we only consider Archimedean copulas with upper tail dependence as described by Theorem 4.3.2 for $\tilde{\theta} \in (1, \infty)$.

4.3.2 Approximation in d_{TV} by a Poisson process

The aim is now to determine an estimate of $d_{TV}(\text{PRM}(\mathbb{E}\Xi_A), \text{PRM}(\boldsymbol{\lambda}))$ for all Archimedean copulas with parameter $\theta \in (1, \infty)$ and $\tilde{\theta} = \theta$, where $\boldsymbol{\lambda}^*$ and λ^* are defined as in (4.23) and (4.22), respectively. We define a function $\bar{\phi}_0$ and its inverse $\bar{\phi}_0^{[-1]}$ as follows:

$$\begin{aligned} \bar{\phi}_0 : [0, 1] &\rightarrow [0, \infty], & r &\mapsto \bar{\phi}_0(r) = r^\theta, \\ \bar{\phi}_0^{[-1]} : [0, \infty] &\rightarrow [0, 1], & x &\mapsto \bar{\phi}_0^{[-1]}(x) = x^{1/\theta}. \end{aligned} \tag{4.27}$$

Now note that we may express λ^* in terms of $\bar{\phi}_0$ and $\bar{\phi}_0^{[-1]}$:

$$\begin{aligned}\lambda^*(s, t) &= \frac{\partial^2}{\partial s \partial t} (-n) \left\{ \left(\frac{s}{n} \right)^\theta + \left(\frac{t}{n} \right)^\theta \right\}^{1/\theta} \\ &= \frac{\partial^2}{\partial s \partial t} (-n) \bar{\phi}_0^{[-1]} \left(\bar{\phi}_0 \left(\frac{s}{n} \right) + \bar{\phi}_0 \left(\frac{t}{n} \right) \right).\end{aligned}$$

The following theorem now gives an upper bound on the error involved in approximating a Poisson process with mean measure $\mathbb{E}\Xi_{A^*}^*$ by another Poisson process with the more useful mean measure λ^* . It uses Proposition 2.5.10 and properties of the generator ϕ . By adding the upper bound from Theorem 4.3.4 below to the error estimate in (4.17), we obtain an upper bound for the total variation distance between $\mathcal{L}(\Xi_{A^*}^*)$ and $\text{PRM}(\lambda^*)$.

Theorem 4.3.4. *Let C_θ be a bivariate Archimedean copula with parameter $\theta \in (1, \infty)$ such that $\tilde{\theta} = \theta$ for $\tilde{\theta}$ defined by (4.24). Let ϕ be the generator of C_θ and suppose that $\bar{\phi}(r) = w^\theta(r) = r^\theta h^\theta(r)$, where w is twice continuously differentiable on $[0, 1]$. Suppose that for some $\delta \in (0, 1)$, we have $h(0) > 0$, $h'(r) \geq 0$, and $w''(r) \geq w''(0)$ for all $r \in [0, \delta]$. For each integer $n \geq 1$, let $(U_1, V_1), \dots, (U_n, V_n)$ be i.i.d. random pairs whose common joint distribution function is given by C_θ . Assume the setting and notation from (4.15)-(4.23), and define $H(r) = \max_{0 \leq \xi \leq r} h'(\xi)$ and $W(r) = \max_{0 \leq \xi \leq r} w''(\xi)$. Then there exists $r_0 \in [0, \delta]$ such that for all $r \leq r_0$, $w'(r) \leq 4h(0)/3$; and for $(s_n, t_n) \in (0, n]^2$ such that*

$$\frac{s_n}{n}, \frac{t_n}{n} \leq \frac{3r_0}{8},$$

it follows that

$$d_{TV}(\text{PRM}(\mathbb{E}\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq \frac{K(s_n + t_n)^2}{n},$$

where $K = K(\theta, h(0), r_0, H(r_0), W(r_0))$ is defined by

$$K = \frac{\pi(\sqrt{2})^\theta}{2} \left[(\theta - 1)\kappa + \left(\frac{4}{3} \right)^{2\theta} \frac{W(r_0)}{h(0)} \right],$$

and

$$\kappa = \frac{H(r_0)}{h(0)} \max \left\{ (\theta + 1) \left(1 + \frac{3r_0 H(r_0)}{16h(0)} \right) \left[1 + \frac{3r_0 H(r_0)}{4h(0)} + \frac{9r_0^2 H(r_0)^2}{256h(0)^2} \right]^\theta, \right. \\ \left. (2\theta - 1)2^{\theta-1} \left(1 + \frac{3r_0 H(r_0)}{8h(0)} \right)^{\theta-1} + 2 \left(1 + \frac{3r_0 H(r_0)}{4h(0)} \right) \right\}.$$

Proof. By Proposition 2.5.10,

$$d_{TV}(\text{PRM}(\mathbb{E}\Xi_{A^*}^*), \text{PRM}(\lambda^*)) \leq \int_0^{t_n} \int_0^{s_n} |e_n^*(s, t) - \lambda^*(s, t)| ds dt,$$

where $e_n^*(s, t)$ and $\lambda^*(s, t)$ are given by (4.20) and (4.22), respectively. Note that

$$\lambda^*(nu, nv) = \frac{1}{n} (\theta - 1)(uv)^{\theta-1} (u^\theta + v^\theta)^{\frac{1}{\theta}-2} \text{ with } (u, v) := (u_n, v_n) := \left(\frac{s}{n}, \frac{t}{n} \right).$$

In order to express $e_n^*(s, t)$ in terms of u and v , note that for $r \in [0, 1]$, and $x = \bar{\phi}(r) = \bar{\phi}_0(w(r))$, with $\bar{\phi}_0$ from (4.27), we have

$$r = \bar{\phi}^{[-1]}(x) = w^{-1}(\bar{\phi}_0^{-1}(x)).$$

Also, note that $\partial^2/(\partial s \partial t) = n^{-2} \partial^2/(\partial u \partial v)$ by the chain rule. Thus, $e_n^*(s, t)$ equals

$$\frac{\partial^2}{\partial s \partial t} (-n) w^{-1} \left[\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{1/\theta} \right] \\ = \left(-\frac{1}{n} \right) \frac{\partial^2}{\partial u \partial v} w^{-1} \left[(w^\theta(u) + w^\theta(v))^{1/\theta} \right],$$

where

$$\frac{\partial^2}{\partial u \partial v} w^{-1} \left[(w^\theta(u) + w^\theta(v))^{1/\theta} \right] \\ = \frac{\{w(u)w(v)\}^{\theta-1} w'(u)w'(v)}{w' \left(w^{-1} \left[(w^\theta(u) + w^\theta(v))^{1/\theta} \right] \right)} (w^\theta(u) + w^\theta(v))^{\frac{1}{\theta}-2} \\ \cdot \left\{ -\frac{w'' \left(w^{-1} \left[(w^\theta(u) + w^\theta(v))^{1/\theta} \right] \right)}{w'^2 \left(w^{-1} \left[(w^\theta(u) + w^\theta(v))^{1/\theta} \right] \right)} (w(u)^\theta + w(v)^\theta)^{1/\theta} + (1 - \theta) \right\}.$$

We may thus bound the integrand $|e_n^*(s, t) - \lambda^*(s, t)| = |e_n^*(nu, nv) - \lambda^*(nu, nv)|$ by

$$\begin{aligned} & \frac{1}{n} \left| \frac{\{w(u)w(v)\}^{\theta-1} \cdot w'(u)w'(v) \cdot w'' \left(w^{-1} \left[(w^\theta(u) + w^\theta(v))^{1/\theta} \right] \right)}{(w^\theta(u) + w^\theta(v))^{2-\frac{2}{\theta}} \cdot w'^3 \left(w^{-1} \left[(w^\theta(u) + w^\theta(v))^{1/\theta} \right] \right)} \right| \\ & + \frac{\theta-1}{n} \left| \frac{\{w(u)w(v)\}^{\theta-1} \cdot w'(u)w'(v) \cdot (w^\theta(u) + w^\theta(v))^{\frac{1}{\theta}-2}}{w' \left(w^{-1} \left[(w^\theta(u) + w^\theta(v))^{1/\theta} \right] \right)} \right. \\ & \quad \left. - (uv)^{\theta-1} (u^\theta + v^\theta)^{\frac{1}{\theta}-2} \right|. \end{aligned} \quad (4.28)$$

Note that since $w(r) = rh(r)$, we have $w'(r) = h(r) + rh'(r)$, and $w''(r) = 2h'(r) + rh''(r)$. Furthermore, $w'(0) = h(0) > 0$, and, since $h'(r) \geq 0$, we have $w''(0) = 2h'(0) \geq 0$ and

$$w'(r) \geq h(0) + rh'(r) \geq h(0), \quad \text{for all } r \in [0, \delta]. \quad (4.29)$$

By continuity of w' at 0, there has to exist $r_0 \in [0, \delta]$ such that

$$w'(r) \leq \frac{4h(0)}{3}, \quad \text{for all } r \leq r_0. \quad (4.30)$$

It follows that

$$h(r) \leq \frac{4h(0)}{3}, \quad \text{for all } r \leq r_0. \quad (4.31)$$

We now determine bounds on the inverse of w close to 0. Since $rh(0) \leq w(r) \leq (4/3)rh(0)$ for all $r \leq r_0$, it follows that for all $x \leq r_0h(0)$,

$$\frac{3x}{4h(0)} \leq w^{-1}(x) \leq \frac{x}{h(0)}. \quad (4.32)$$

See Figure 4.6 for an illustration. Note that with the well known property

$$2^{\frac{1}{p}-\frac{1}{q}} \|(s, t)\|_q \leq \|(s, t)\|_p \leq \|(s, t)\|_q$$

of the p -norm $\|(s, t)\|_p := (s^p + t^p)^{1/p}$ that holds if $p > q > 0$, we obtain the following inequalities:

$$(s^\theta + t^\theta)^{\frac{1}{\theta}} \leq s + t, \quad s^{\theta+1} + t^{\theta+1} \leq (s + t)^{\theta+1}, \quad s^\theta + t^\theta \geq 2^{1-\theta}(s + t)^\theta. \quad (4.33)$$

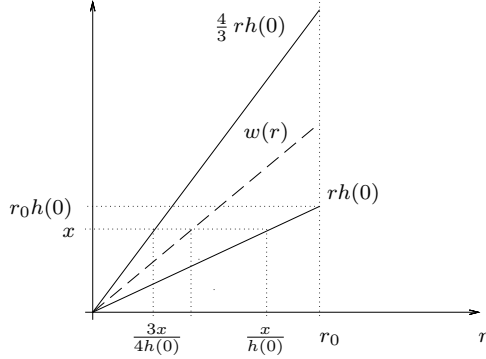


Figure 4.6: If $w(r)$ is wedged between two straight lines for all $r \leq r_0$, then its inverse must be wedged between the inverses of these lines, i.e. $\frac{3x}{4h(0)} \leq w^{-1}(x) \leq \frac{x}{h(0)}$, for all $x \leq r_0 h(0)$.

For $(s, t) \in (0, s_n] \times (0, t_n] \subseteq (0, n]^2$ such that

$$u = \frac{s}{n} \leq \frac{s_n}{n} \leq \frac{3r_0}{8} \quad \text{and} \quad v = \frac{t}{n} \leq \frac{t_n}{n} \leq \frac{3r_0}{8},$$

we thus have, using (4.33) and (4.31),

$$\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \leq w \left(\frac{s}{n} \right) + w \left(\frac{t}{n} \right) \leq \frac{4h(0)(s+t)}{3n} \leq r_0 h(0).$$

Therefore, by (4.32),

$$\begin{aligned} 0 &\leq \frac{3}{4h(0)} \left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \\ &\leq w^{-1} \left[\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \right] \leq \frac{1}{h(0)} \left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \leq r_0. \end{aligned} \tag{4.34}$$

We proceed by determining bounds on the first of the two error terms in (4.28). By (4.31), and since $h(r) \geq h(0)$ (due to $h'(r) \geq 0$ for all $r \in [0, \delta)$),

$$\left(\frac{h(0)}{n}\right)^{2\theta-2} (st)^{\theta-1} \leq \left\{w\left(\frac{s}{n}\right) w\left(\frac{t}{n}\right)\right\}^{\theta-1} \leq \left(\frac{4}{3}\right)^{2\theta-2} \left(\frac{h(0)}{n}\right)^{2\theta-2} (st)^{\theta-1}. \quad (4.35)$$

Moreover, since $\frac{2}{\theta} - 2 < 0$, we obtain

$$\begin{aligned} \left(\frac{4}{3}\right)^{2-2\theta} \left(\frac{h(0)}{n}\right)^{2-2\theta} (s^\theta + t^\theta)^{\frac{2}{\theta}-2} \\ \leq \left(w^\theta\left(\frac{s}{n}\right) + w^\theta\left(\frac{t}{n}\right)\right)^{\frac{2}{\theta}-2} \leq \left(\frac{h(0)}{n}\right)^{2-2\theta} (s^\theta + t^\theta)^{\frac{2}{\theta}-2}. \end{aligned} \quad (4.36)$$

By (4.29) and (4.30),

$$h(0)^2 \leq w'\left(\frac{s}{n}\right) w'\left(\frac{t}{n}\right) \leq \left(\frac{4}{3}\right)^2 h(0)^2, \quad (4.37)$$

and also, due to (4.34),

$$h(0) \leq w' \left(w^{-1} \left[\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \right] \right) \leq \frac{4}{3} h(0).$$

It follows that

$$\left(\frac{3}{4h(0)}\right)^3 \leq \frac{1}{w'^3 \left(w^{-1} \left[\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \right] \right)} \leq \frac{1}{h(0)^3}. \quad (4.38)$$

Furthermore, by (4.34), and since $w''(r) \geq w''(0) = 2h'(0)$,

$$0 \leq 2h'(0) \leq w'' \left(w^{-1} \left[\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \right] \right) \leq \max_{0 \leq \xi \leq r_0} w''(\xi) = W(r_0). \quad (4.39)$$

Inequalities (4.35)-(4.39) yield the following upper bound on the first of the

two error terms in (4.28):

$$\begin{aligned}
 & \frac{1}{n} \cdot \frac{\left\{w\left(\frac{s}{n}\right) w\left(\frac{t}{n}\right)\right\}^{\theta-1} w'\left(\frac{s}{n}\right) w'\left(\frac{t}{n}\right) w''\left(w^{-1}\left[\left(w^{\theta}\left(\frac{s}{n}\right)+w^{\theta}\left(\frac{t}{n}\right)\right)^{1 / \theta}\right]\right)}{\left(w^{\theta}\left(\frac{s}{n}\right)+w^{\theta}\left(\frac{t}{n}\right)\right)^{2-\frac{2}{\theta}} w'^3\left(w^{-1}\left[\left(w^{\theta}\left(\frac{s}{n}\right)+w^{\theta}\left(\frac{t}{n}\right)\right)^{1 / \theta}\right]\right)} (\geq 0) \\
 & \leq \frac{1}{n}\left(\frac{4}{3}\right)^{2 \theta} \frac{W\left(r_0\right)}{h(0)}(s t)^{\theta-1}\left(s^{\theta}+t^{\theta}\right)^{\frac{2}{\theta}-2} \\
 & \leq \frac{s+t}{n}\left(\frac{4}{3}\right)^{2 \theta} \frac{W\left(r_0\right)}{h(0)}(s t)^{\theta-1}\left(s^{\theta}+t^{\theta}\right)^{\frac{1}{\theta}-2},
 \end{aligned} \tag{4.40}$$

where we used (4.33) for the last inequality. We next determine bounds on the second of the two error terms in (4.28). Note that for all $r \in[0, \delta]$,

$$h(0) \leq h(r) \leq h(0)+r \max _{0 \leq \xi \leq r} h'(r)=h(0)+r H(r), \tag{4.41}$$

and thereby,

$$h(0) \leq h(0)+r h'(r) \leq w'(r) \leq h(0)+r H(r)+r h'(r) \leq h(0)+2 r H(r) . \tag{4.42}$$

By (4.34) and (4.42),

$$\begin{aligned}
 h(0) & \leq w'\left(w^{-1}\left[\left(w^{\theta}\left(\frac{s}{n}\right)+w^{\theta}\left(\frac{t}{n}\right)\right)^{\frac{1}{\theta}}\right]\right) \\
 & \leq h(0)+2 w^{-1}\left[\left(w^{\theta}\left(\frac{s}{n}\right)+w^{\theta}\left(\frac{t}{n}\right)\right)^{\frac{1}{\theta}}\right] \\
 & \quad \cdot H\left(w^{-1}\left[\left(w^{\theta}\left(\frac{s}{n}\right)+w^{\theta}\left(\frac{t}{n}\right)\right)^{\frac{1}{\theta}}\right]\right),
 \end{aligned}$$

and by (4.34), (4.41) and (4.33),

$$\begin{aligned}
 w^{-1}\left[\left(w^{\theta}\left(\frac{s}{n}\right)+w^{\theta}\left(\frac{t}{n}\right)\right)^{\frac{1}{\theta}}\right] & \leq \frac{1}{h(0)}\left(w^{\theta}\left(\frac{s}{n}\right)+w^{\theta}\left(\frac{t}{n}\right)\right)^{\frac{1}{\theta}} \\
 & \leq \frac{\frac{s}{n} h\left(\frac{s}{n}\right)+\frac{t}{n} h\left(\frac{t}{n}\right)}{h(0)} \\
 & \leq \frac{s+t}{n}+\frac{s^2 H\left(\frac{s}{n}\right)+t^2 H\left(\frac{t}{n}\right)}{n^2 h(0)} .
 \end{aligned}$$

Furthermore, note that since

$$w^{-1} \left[\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \right] \leq r_0 \quad \text{and} \quad \frac{s}{n}, \frac{t}{n} \leq \frac{3r_0}{8},$$

we have

$$H \left(w^{-1} \left[\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \right] \right) \leq H(r_0) \quad \text{and} \quad H \left(\frac{s}{n} \right), H \left(\frac{t}{n} \right) \leq H(r_0). \quad (4.43)$$

It follows that

$$\begin{aligned} h(0) &\leq w' \left(w^{-1} \left[\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \right] \right) \\ &\leq h(0) + 2H(r_0) \frac{s+t}{n} + \frac{2H^2(r_0)}{h(0)} \left(\frac{s+t}{n} \right)^2, \end{aligned}$$

and thereby

$$\begin{aligned} \frac{1}{h(0)} \left\{ 1 - \frac{2H(r_0)(s+t)}{h(0)n} - \frac{2H^2(r_0)}{h(0)^2} \left(\frac{s+t}{n} \right)^2 \right\} \\ \leq \frac{1}{w' \left(w^{-1} \left[\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}} \right] \right)} \leq \frac{1}{h(0)}, \quad (4.44) \end{aligned}$$

where we used for the lower bound that $(b+az)^{-1} \geq b^{-1} - azb^{-2}$ for $b > 0$, $a \geq 0$ and $z < 1$. Similarly to (4.36), we have

$$\left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}-2} \leq \left(\frac{h(0)}{n} \right)^{1-2\theta} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} \quad (4.45)$$

For a lower bound, we successively use (4.41), (4.43), the inequalities $(1+z)^\theta \leq 1 + \theta z(1+z)^{\theta-1}$ and $(1+z)^{1/\theta-2} \geq 1 - (2-1/\theta)z$, for $0 \leq z \leq 1$, as well as

$s/n, t/n \leq 3r_0/8$ and (4.33). That is,

$$\begin{aligned}
 & \left(w^\theta \left(\frac{s}{n} \right) + w^\theta \left(\frac{t}{n} \right) \right)^{\frac{1}{\theta}-2} \\
 & \geq \left(\frac{h(0)}{n} \right)^{1-2\theta} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} \left\{ \frac{s^\theta \left[1 + \frac{sH(r_0)}{nh(0)} \right]^\theta + t^\theta \left[1 + \frac{tH(r_0)}{nh(0)} \right]^\theta}{s^\theta + t^\theta} \right\}^{\frac{1}{\theta}-2} \\
 & \geq \left(\frac{h(0)}{n} \right)^{1-2\theta} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} \left\{ 1 + \frac{\theta H(r_0) (s^{\theta+1} + t^{\theta+1})}{nh(0)(s^\theta + t^\theta)} \right. \\
 & \quad \cdot \left. \left[1 + \frac{3r_0 H(r_0)}{8h(0)} \right]^{\theta-1} \right\}^{\frac{1}{\theta}-2} \\
 & \geq \left(\frac{h(0)}{n} \right)^{1-2\theta} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} \left\{ 1 - \frac{(2\theta-1)2^{\theta-1}H(r_0)(s+t)}{nh(0)} \right. \\
 & \quad \cdot \left. \left[1 + \frac{3r_0 H(r_0)}{8h(0)} \right]^{\theta-1} \right\}. \tag{4.46}
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 & \left(\frac{h(0)}{n} \right)^{2\theta-2} (st)^{\theta-1} \leq \left\{ w \left(\frac{s}{n} \right) w \left(\frac{t}{n} \right) \right\}^{\theta-1} \\
 & \leq \left(\frac{h(0)}{n} \right)^{2\theta-2} (st)^{\theta-1} \left[1 + \frac{sH(r_0)}{nh(0)} \right]^{\theta-1} \left[1 + \frac{tH(r_0)}{nh(0)} \right]^{\theta-1}. \tag{4.47}
 \end{aligned}$$

By (4.42) and (4.43),

$$h(0)^2 \leq w' \left(\frac{s}{n} \right) w' \left(\frac{t}{n} \right) \leq h(0)^2 \left[1 + \frac{2sH(r_0)}{nh(0)} \right] \left[1 + \frac{2tH(r_0)}{nh(0)} \right]. \tag{4.48}$$

Multiplication of (4.47) and (4.48) gives

$$\begin{aligned}
 & \left(\frac{1}{n}\right)^{2\theta-2} h(0)^{2\theta} (st)^{\theta-1} \\
 & \leq \left\{ w\left(\frac{s}{n}\right) w\left(\frac{t}{n}\right) \right\}^{\theta-1} w'\left(\frac{s}{n}\right) w'\left(\frac{t}{n}\right) \\
 & \leq \left(\frac{1}{n}\right)^{2\theta-2} h(0)^{2\theta} (st)^{\theta-1} \left[1 + \frac{sH(r_0)}{nh(0)}\right]^{\theta+1} \left[1 + \frac{tH(r_0)}{nh(0)}\right]^{\theta+1} \\
 & \leq \left(\frac{1}{n}\right)^{2\theta-2} h(0)^{2\theta} (st)^{\theta-1} \left[1 + \frac{H(r_0)(s+t)}{nh(0)} + \frac{H(r_0)^2(s+t)^2}{4n^2h(0)^2}\right]^{\theta+1},
 \end{aligned} \tag{4.49}$$

where we used $(1+2z) \leq (1+z)^2$ for the second inequality, and $st \leq (s+t)^2/4$ (due to the inequality of arithmetic and geometric means) for the third. For the part in brackets, we use $(s+t)/n \leq 3r_0/4$ and the inequality $(1+z)^{\theta+1} \leq 1 + (\theta+1)z(1+z)^\theta$. We obtain

$$\begin{aligned}
 & \left[1 + \frac{H(r_0)(s+t)}{nh(0)} + \frac{H(r_0)^2(s+t)^2}{4n^2h(0)^2}\right]^{\theta+1} \\
 & \leq \left[1 + \frac{H(r_0)(s+t)}{nh(0)} \left(1 + \frac{3r_0H(r_0)}{16h(0)}\right)\right]^{\theta+1} \\
 & \leq 1 + \frac{(\theta+1)H(r_0)(s+t)}{nh(0)} \left(1 + \frac{3r_0H(r_0)}{16h(0)}\right) \left[1 + \frac{3r_0H(r_0)}{4h(0)} + \frac{9r_0^2H(r_0)^2}{256h(0)^2}\right]^\theta.
 \end{aligned} \tag{4.50}$$

By (4.44)-(4.50), we obtain, on the one hand,

$$\begin{aligned}
 & \frac{\left\{ w\left(\frac{s}{n}\right) w\left(\frac{t}{n}\right) \right\}^{\theta-1} \cdot w'\left(\frac{s}{n}\right) w'\left(\frac{t}{n}\right) \cdot \left(w^\theta\left(\frac{s}{n}\right) + w^\theta\left(\frac{t}{n}\right)\right)^{\frac{1}{\theta}-2}}{w'\left(w^{-1}\left[\left(w^\theta\left(\frac{s}{n}\right) + w^\theta\left(\frac{t}{n}\right)\right)^{1/\theta}\right]\right)} \\
 & \quad - n(st)^{\theta-1} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} \\
 & \leq (st)^{\theta-1} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} (s+t) \\
 & \quad \cdot \frac{(\theta+1)H(r_0)}{h(0)} \left(1 + \frac{3r_0H(r_0)}{16h(0)}\right) \left[1 + \frac{3r_0H(r_0)}{4h(0)} + \frac{9r_0^2H(r_0)^2}{256h(0)^2}\right]^\theta,
 \end{aligned}$$

and, on the other hand, again using $(s+t)/n \leq 3r_0/4$ for the lower bound in (4.44),

$$\begin{aligned}
 & n(st)^{\theta-1} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} \\
 & \quad - \frac{\left\{w\left(\frac{s}{n}\right) w\left(\frac{t}{n}\right)\right\}^{\theta-1} \cdot w'\left(\frac{s}{n}\right) w'\left(\frac{t}{n}\right) \cdot \left(w^\theta\left(\frac{s}{n}\right) + w^\theta\left(\frac{t}{n}\right)\right)^{\frac{1}{\theta}-2}}{w'\left(w^{-1}\left[\left(w^\theta\left(\frac{s}{n}\right) + w^\theta\left(\frac{t}{n}\right)\right)^{1/\theta}\right]\right)} \\
 & \leq n(st)^{\theta-1} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} \\
 & \quad \cdot \left\{1 - \left[1 - \frac{(2\theta-1)2^{\theta-1}H(r_0)(s+t)}{nh(0)} \left(1 + \frac{3r_0H(r_0)}{8h(0)}\right)^{\theta-1}\right] \right. \\
 & \quad \quad \left. \cdot \left[1 - \frac{2H(r_0)(s+t)}{h(0)n} \left(1 + \frac{3r_0H(r_0)}{4h(0)}\right)\right]\right\} \\
 & \leq (st)^{\theta-1} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} (s+t) \\
 & \quad \cdot \frac{H(r_0)}{h(0)} \left[(2\theta-1)2^{\theta-1} \left(1 + \frac{3r_0H(r_0)}{8h(0)}\right)^{\theta-1} + 2 \left(1 + \frac{3r_0H(r_0)}{4h(0)}\right) \right].
 \end{aligned}$$

Hence, the second of the two error terms in (4.28) is bounded by

$$(\theta-1)(st)^{\theta-1} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} \frac{\kappa(s+t)}{n}, \quad (4.51)$$

where

$$\begin{aligned}
 \kappa = \frac{H(r_0)}{h(0)} \max & \left\{ (\theta+1) \left(1 + \frac{3r_0H(r_0)}{16h(0)}\right) \left[1 + \frac{3r_0H(r_0)}{4h(0)} + \frac{9r_0^2H(r_0)^2}{256h(0)^2}\right]^\theta, \right. \\
 & \left. (2\theta-1)2^{\theta-1} \left(1 + \frac{3r_0H(r_0)}{8h(0)}\right)^{\theta-1} + 2 \left(1 + \frac{3r_0H(r_0)}{4h(0)}\right) \right\}.
 \end{aligned}$$

Thus, by (4.40) and (4.51),

$$\begin{aligned}
 & \int_0^{t_n} \int_0^{s_n} |e_n^*(s, t) - \lambda^*(s, t)| \, ds dt \\
 & \leq \int_0^{t_n} \int_0^{s_n} \lambda^*(s, t) \cdot \frac{s+t}{n} \cdot \left[\kappa + \left(\frac{4}{3} \right)^{2\theta} \frac{W(r_0)}{h(0)(\theta-1)} \right] \\
 & \leq \lambda^*(A^*) \cdot \frac{s_n + t_n}{n} \cdot \left[\kappa + \left(\frac{4}{3} \right)^{2\theta} \frac{W(r_0)}{h(0)(\theta-1)} \right],
 \end{aligned}$$

where we used the mean value theorem for integration for the second inequality. Using polar coordinates $s = \rho \cos \varphi > 0$ and $t = \rho \sin \varphi > 0$, with $\rho = \sqrt{s^2 + t^2}$ and $\varphi \in (0, \pi/2)$, we can rewrite $\lambda^*(A^*)$ as follows:

$$\begin{aligned}
 & \int_0^{s_n} \int_0^{t_n} (\theta-1)(st)^{\theta-1} (s^\theta + t^\theta)^{\frac{1}{\theta}-2} \, dt ds \\
 & \leq \int_0^{\sqrt{s_n^2 + t_n^2}} \int_0^{\frac{\pi}{2}} (\theta-1)(\cos \varphi \sin \varphi)^{\theta-1} (\cos^\theta \varphi + \sin^\theta \varphi)^{\frac{1}{\theta}-2} \, d\varphi d\rho \\
 & = (\theta-1) \sqrt{s_n^2 + t_n^2} \int_0^{\frac{\pi}{2}} (\cos \varphi \sin \varphi)^{\theta-1} (\cos^\theta \varphi + \sin^\theta \varphi)^{\frac{1}{\theta}-2} \, d\varphi \\
 & \leq \frac{\pi}{2} (\theta-1) \sqrt{s_n^2 + t_n^2} \cdot \max_{\varphi \in (0, \frac{\pi}{2})} \left| (\cos \varphi \sin \varphi)^{\theta-1} (\cos^\theta \varphi + \sin^\theta \varphi)^{\frac{1}{\theta}-2} \right|.
 \end{aligned}$$

Note that for all $\varphi \in (0, \pi/4]$, we have $(\sqrt{2})^{-1} \leq \cos \varphi < 1$ and $0 < \sin \varphi \leq (\sqrt{2})^{-1}$. Therefore, $\cos^\theta \varphi + \sin^\theta \varphi > (\sqrt{2})^{-\theta}$ and $\cos \varphi \sin \varphi < (\sqrt{2})^{-1}$. Analogously, we get the same inequalities for all $\varphi \in [\pi/4, \pi/2)$. Thus, we obtain

$$(\cos^\theta \varphi + \sin^\theta \varphi)^{\frac{1}{\theta}-2} < (\sqrt{2})^{2\theta-1} \quad \text{and} \quad (\cos \varphi \sin \varphi)^{\theta-1} < (\sqrt{2})^{1-\theta}$$

for all $\varphi \in (0, \pi/2)$, and therefore

$$\lambda^*(A^*) \leq \frac{\pi}{2} (\theta-1) (\sqrt{2})^\theta \sqrt{s_n^2 + t_n^2} \leq \frac{\pi}{2} (\theta-1) (\sqrt{2})^\theta (s_n + t_n). \quad (4.52)$$

Hence,

$$\begin{aligned} & \int_0^{t_n} \int_0^{s_n} |e_n^*(s, t) - \lambda^*(s, t)| \, ds dt \\ & \leq \frac{(s_n + t_n)^2}{n} \cdot \frac{\pi(\sqrt{2})^\theta}{2} \cdot \left[(\theta - 1)\kappa + \left(\frac{4}{3}\right)^{2\theta} \frac{W(r_0)}{h(0)} \right]. \end{aligned}$$

□

For any choice of Archimedean copula satisfying (4.24), we thus first need to determine $h(0)$ and w' by way of the generator ϕ , in order to find r_0 such that for all $r \leq r_0$, it holds that $w'(r) \leq 4h(0)/3$. We can then choose n , s_n and t_n such that

$$\frac{s_n}{n}, \frac{t_n}{n} \leq \frac{3r_0}{8}. \quad (4.53)$$

Suppose, for instance, that this is satisfied for $s_n = t_n = \sqrt{\log n}/2$ for some big enough integer n . For an MPPE whose bivariate marks are distributed according to the chosen copula, the error of the approximation in total variation by a Poisson process with intensity measure λ^* is then, by (4.17) and Theorem 4.3.4, bounded by

$$\frac{\sqrt{\log n}}{2n} + \frac{K \log n}{n}, \quad (4.54)$$

and n needs to be big enough to offset the effect of the multiplication by the constant K . The expected number of exceedances $\mathbb{E}W_{A^*}^*$ of the MPPE is then approximately

$$\lambda^*(A^*) = s_n + t_n - (s_n^\theta + t_n^\theta)^{\frac{1}{\theta}} = \left(1 - 2^{\frac{1}{\theta}-1}\right) \log n. \quad (4.55)$$

The next section gives some examples.

4.3.3 Examples

We apply Theorem 4.3.4 to several examples. Charpentier and Segers (2009) showed that (4.24) and thereby (4.25) are satisfied for the families of Archimedean copulas listed in Table 4.1, with $\tilde{\theta} = \theta \in [1, \infty)$. For each of these, we assume that $\theta > 1$ (as the case $\theta = 1$ gives independence in the upper tail). We denote the various families of copulas by the numbers assigned to them in Nelsen (2006) and Charpentier and Segers (2009).

Note that $\bar{\phi}$ of family (2) is exactly $\bar{\phi}_0$ from (4.27). The intensity measure of the corresponding MPPE thus equals the intensity measure $\boldsymbol{\lambda}^*$ of the Poisson process that we would approximate by in Theorem 4.3.4. For family (2), (4.17) is thus sufficient, and there is no need to apply Theorem 4.3.4. Note moreover that family (4) is the family of Gumbel copulas from Example 4.3.1.

Nr.	$C_\theta(u, v)$	$\bar{\phi}(r) = \phi(1 - r)$
(2)	$\max \left\{ 1 - [(1 - u)^\theta + (1 - v)^\theta]^{\frac{1}{\theta}}, 0 \right\}$	r^θ
(4)	$\exp \left\{ - [(-\log u)^\theta + (-\log v)^\theta]^{\frac{1}{\theta}} \right\}$	$[-\log(1 - r)]^\theta$
(6)	$1 - [(1 - u)^\theta + (1 - v)^\theta - (1 - u)^\theta(1 - v)^\theta]^{\frac{1}{\theta}}$	$-\log(1 - r^\theta)$
(12)	$\left\{ 1 + [(u^{-1} - 1)^\theta + (v^{-1} - 1)^\theta]^{\frac{1}{\theta}} \right\}^{-1}$	$\left(\frac{r}{1 - r} \right)^\theta$
(14)	$\left\{ 1 + \left[(u^{-\frac{1}{\theta}} - 1)^\theta + (v^{-\frac{1}{\theta}} - 1)^\theta \right]^{\frac{1}{\theta}} \right\}^{-\theta}$	$[(1 - r)^{-\frac{1}{\theta}} - 1]^\theta$
(15)	$\left\{ \max \left(1 - \left[\left(1 - u^{\frac{1}{\theta}} \right)^\theta + \left(1 - v^{\frac{1}{\theta}} \right)^\theta \right]^{\frac{1}{\theta}}, 0 \right) \right\}^\theta$	$\left[1 - (1 - r)^{\frac{1}{\theta}} \right]^\theta$
(21)	$1 - (1 - \{ \max([1 - (1 - u)^\theta]^{\frac{1}{\theta}} + [1 - (1 - v)^\theta]^{\frac{1}{\theta}} - 1, 0) \}^\theta)^{\frac{1}{\theta}}$	$1 - (1 - r^\theta)^{\frac{1}{\theta}}$

Table 4.1: We list families of Archimedean copulas with parameter $\theta \in [1, \infty)$, for which the limit $\tilde{\theta}$ in (4.24) exists in $[1, \infty)$ and equals θ . These copulas exhibit upper tail dependence as determined by Theorem 4.3.2 (unless $\tilde{\theta} = \theta = 1$ in which case we have asymptotic independence).

For each of the examples from Table 4.1 (with $\theta > 1$), it is possible to show that the function $w(r) = \bar{\phi}^{\frac{1}{\theta}}(r)$ is twice continuously differentiable on $[0, 1)$.

Consider, e.g., family (4), for which $w(r) = -\log(1-r)$:

$$w'(r) = \frac{1}{1-r}, \quad w'(0) = 1, \quad w''(r) = \frac{1}{(1-r)^2}, \quad w''(0) = 1.$$

For each example, we now give the function $h(r) = \bar{\phi}^{\frac{1}{\theta}}(r)/r$ and indicate the first few terms of its series expansion. We thereby determine the value $h(0)$.

(2): $h(r) \equiv 1$. We have $h(0) = 1$.

(4): Using series expansion of the logarithm, we have

$$h(r) = -\frac{\log(1-r)}{r} = \frac{r + \frac{r^2}{2} + \frac{r^3}{3} + \dots}{r} = 1 + \frac{r}{2} + \frac{r^2}{3} + \dots,$$

and thus $h(0) = 1$.

(6): Series expansion of the logarithm gives

$$h(r) = \frac{[-\log(1-r^\theta)]^{1/\theta}}{r} = \left(\sum_{j=1}^{\infty} \frac{r^{(j-1)\theta}}{j} \right)^{\frac{1}{\theta}} = \left(1 + \sum_{j=2}^{\infty} \frac{r^{(j-1)\theta}}{j} \right)^{\frac{1}{\theta}}.$$

Taylor expansion about 0 gives

$$(1+z)^{\frac{1}{\theta}} = 1 + \frac{1}{\theta} z + \frac{1}{\theta} \left(\frac{1}{\theta} - 1 \right) \frac{z^2}{2} + \dots,$$

and with $z = \sum_{j=2}^{\infty} \frac{r^{(j-1)\theta}}{j}$, we obtain

$$\begin{aligned} h(r) &= 1 + \frac{1}{2\theta} r^\theta + \frac{1}{8\theta} \left(\frac{1}{\theta} + \frac{5}{3} \right) r^{2\theta} + \frac{1}{6\theta} \left(\frac{1}{\theta} + \frac{1}{2} \right) r^{3\theta} + \dots, \\ h'(r) &= \frac{1}{2} r^{\theta-1} + \frac{1}{4} \left(\frac{1}{\theta} + \frac{5}{3} \right) r^{2\theta-1} + \frac{1}{2} \left(\frac{1}{\theta} + \frac{1}{2} \right) r^{3\theta-1} + \dots, \\ h''(r) &= \frac{\theta-1}{2} r^{\theta-2} + \frac{2\theta-1}{4} \left(\frac{1}{\theta} + \frac{5}{3} \right) r^{2\theta-2} + \frac{3\theta-1}{2} \left(\frac{1}{\theta} + \frac{1}{2} \right) r^{3\theta-2} + \dots \end{aligned}$$

Thus $h(0) = 1$. Note that h is not twice differentiable at 0 if $\theta < 2$; however, Theorem 4.3.4 uses only $w''(r) = h'(r) + rh''(r)$, which does exist at $r = 0$ for all $\theta > 1$. Family (21) below behaves similarly.

(12): The function

$$h(r) = \frac{1}{1-r} = \sum_{j=0}^{\infty} r^j = 1 + r + r^2 + \dots$$

is the geometric series. We have $h(0) = 1$.

(14): Taylor expansion about 0 gives

$$(1-r)^{-\frac{1}{\theta}} = 1 + \frac{1}{\theta} r + \frac{1}{2\theta} \left(\frac{1}{\theta} + 1 \right) r^2 + \dots,$$

and therefore,

$$h(r) = \frac{(1-r)^{-1/\theta} - 1}{r} = \frac{1}{\theta} + \frac{1}{2\theta} \left(\frac{1}{\theta} + 1 \right) r + \dots$$

We thus obtain $h(0) = \frac{1}{\theta}$.

(15): With Taylor expansion about 0, we find

$$(1-r)^{\frac{1}{\theta}} = 1 - \frac{1}{\theta} r - \frac{1}{2\theta} \left(1 - \frac{1}{\theta} \right) r^2 - \frac{1}{6\theta} \left(1 - \frac{1}{\theta} \right) \left(2 - \frac{1}{\theta} \right) r^3 - \dots, \quad (4.56)$$

and therefore,

$$h(r) = \frac{1 - (1-r)^{1/\theta}}{r} = \frac{1}{\theta} + \frac{1}{2\theta} \left(1 - \frac{1}{\theta} \right) r + \dots,$$

and $h(0) = \frac{1}{\theta}$.

(21): By (4.56), we find

$$\begin{aligned} h(r) &= \frac{\left[1 - (1-r^\theta)^{1/\theta} \right]^{1/\theta}}{r} \\ &= \left\{ \frac{1}{\theta} + \frac{1}{2\theta} \left(1 - \frac{1}{\theta} \right) r^\theta + \frac{1}{6\theta} \left(1 - \frac{1}{\theta} \right) \left(2 - \frac{1}{\theta} \right) r^{2\theta} + \dots \right\}^{\frac{1}{\theta}}, \end{aligned}$$

and Taylor expansion about 0 gives

$$\left(\frac{1}{\theta} + z\right)^{\frac{1}{\theta}} = \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}} + \frac{1}{\theta} \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}-1} z - \frac{1}{2\theta} \left(1 - \frac{1}{\theta}\right) \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}-2} z^2 + \dots$$

Then,

$$h(r) = \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}} + \frac{1}{2\theta^2} \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}-1} \left(1 - \frac{1}{\theta}\right) r^{\theta} + \dots \quad \text{and} \quad h(0) = \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}}.$$

We have thus shown that $h(0) > 0$ for each of the examples. Note that it is furthermore possible to show that $h'(r) \geq 0$ and $w''(r) \geq w''(0)$ for all r close to 0 (or even for all $r \in [0, 1)$); see also Figures 4.7 and 4.8. We next need to compute the values

$$r_0, \quad H(r_0) = \max_{0 \leq \xi \leq r_0} h'(\xi), \quad W(r_0) = \max_{0 \leq \xi \leq r_0} w''(\xi),$$

in order to determine the constant K . Note that for families (6), (14), (15) and (21), the functions h and w depend on the parameter θ and the value of θ should thus be specified. For families (4) and (12), this is not the case. However, we still need to specify θ for all families of copulas, as the constant K depends on θ . By way of example, we compute r_0 , $H(r_0)$, $W(r_0)$ and finally $K = K(\theta, h(0), r_0, H(r_0), W(r_0))$ for each family of copulas for the parameter values $\theta = 1.5$ and $\theta = 3$, respectively. The results are summarised in Tables 4.2 and 4.3, respectively. Figures 4.7 and 4.8 illustrate that for $\theta = 1.5$, the functions h' and w'' are non-decreasing, and that therefore $H(r_0) = h'(r_0)$ and $W(r_0) = w''(r_0)$. The functions h' and w'' behave analogously for $\theta = 3$.

Nr.	$h(0)$	r_0	$H(r_0)$	$W(r_0)$	K
(4)	1	0.250	0.731	1.778	16.2
(6)	1	0.851	2.531	21.027	186.0
(12)	1	0.133	1.331	3.080	28.4
(14)	2/3	0.158	0.754	1.761	24.3
(15)	2/3	0.578	0.229	0.703	9.0
(21)	$(2/3)^{\frac{2}{3}}$	0.738	0.240	1.053	10.8

Table 4.2: For $\theta = 1.5$, we compute the values $h(0)$, r_0 , $H(r_0)$, $W(r_0)$ and the constant K for each of the examples of families of copulas from Table 4.1.

Remark 4.3.5. Note that $h(0)$ does not depend on the value of θ . We thus do not include it again in Table 4.3 for $\theta = 3$. Also note that in Tables 4.2 and 4.3, r_0 was rounded down to three decimal places (any r smaller than r_0 would indeed do), but that we used six decimal places to compute $H(r_0)$ and $W(r_0)$. In Tables 4.2 and 4.3, the results for $H(r_0)$ and $W(r_0)$ are rounded to three decimal places, and the results for K to one. As can be seen when considering the formula for K in Theorem 4.3.4 or when comparing the two tables, the value of K increases as θ increases.

The results from the tables can be interpreted by considerations like the following: for instance, if we choose the Gumbel copula, i.e. copula (4), with

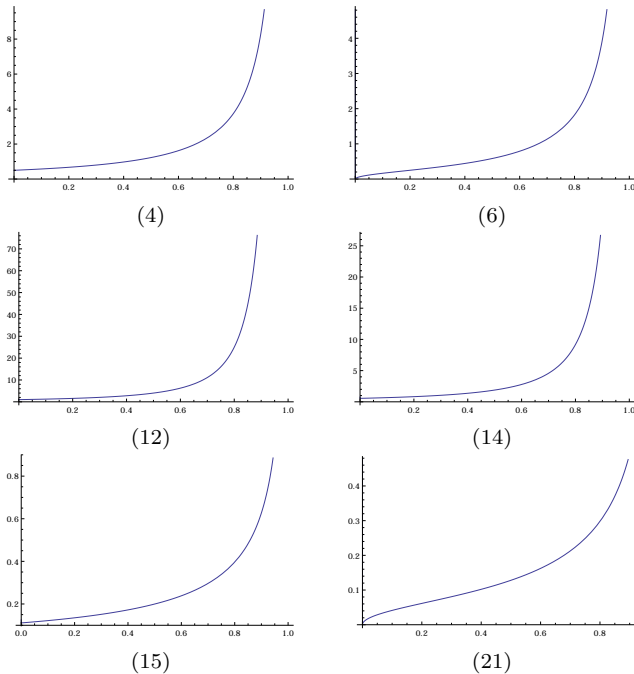


Figure 4.7: Plots of h' on $[0, 1]$ for each of the families of copulas from Table 4.1, for $\theta = 1.5$.

$\theta = 1.5$, and if we choose $s_n = t_n = \sqrt{\log n}/2$, the sample size n needs to be big enough to allow for

$$\frac{\sqrt{\log n}}{2n} \leq \frac{3r_0}{8} = \frac{3}{32} \approx 0.0938.$$

This is satisfied for each integer $n \geq 8$. As determined in (4.55), the expected number of exceedances is approximately

$$\left(1 - 2^{-\frac{1}{3}}\right) \sqrt{\log n} \approx 0.2\sqrt{\log n},$$

and the overall error bound from (4.54) with $K = 16.2$ is smaller than 1 only for $n \geq 70$. For example, for $n = 100$, we only expect $0.2\sqrt{\log 100} \approx 0.43$, i.e. less than one joint exceedance of the thresholds. The sample size thus has to be very big in order to expect only as much as one threshold exceedance and to get a small error.

Nr.	r_0	$H(r_0)$	$W(r_0)$	K
(4)	0.250	0.731	1.778	207.2
(6)	0.701	0.375	2.078	1401.1
(12)	0.133	1.331	3.080	372.4
(14)	0.194	0.291	0.736	313.9
(15)	0.350	1.773	0.457	107.3
(21)	0.774	0.238	1.479	126.1

Table 4.3: For $\theta = 3$, we compute the values $h(0)$, r_0 , $H(r_0)$, $W(r_0)$ and the constant K for each of the families of copulas from Table 4.1.

Some values for K in the tables might be unnecessarily high. These might be reduced by choosing a smaller r_0 than the one indicated. For instance, for copula (6) with $\theta = 3$, Table 4.3 indicates $r_0 = 0.701$. We might instead choose, say, $r_0 = 0.1$, as the inequality $w'(r) \leq 4h(0)/3$ will then still be satisfied. Then, $H(r_0) = 0.005$, $W(r_0) = 0.02$ and $K = 1.5$ (instead of $K = 1401.1$ as before). Condition (4.53) is satisfied for s_n and t_n chosen as above for all integers $n \geq 24$. For $n = 24$, the size of the error bound is ≈ 0.24 and the expected number of joint threshold exceedances is ≈ 0.63 , showing again that an even bigger sample size is needed to expect at least one joint threshold exceedance.

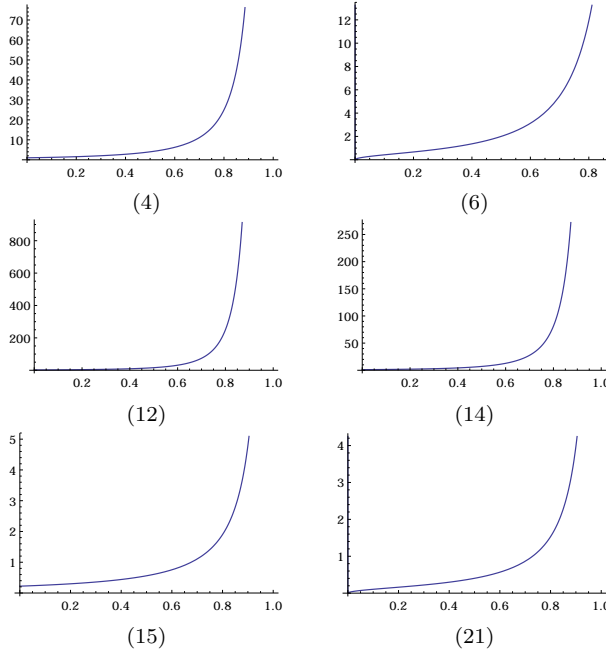


Figure 4.8: Plots of w'' on $[0, 1]$ for each of the families of copulas from Table 4.1, for $\theta = 1.5$.

4.3.4 Higher dimensions

Sections 4.3.1-4.3.3 treat the case $d = 2$. Theorem 4.3.2 is, however, also valid for $d \geq 3$; see Theorem 4.1 in Charpentier and Segers (2009). This section sketches two ways to determine bounds on the error between $\mathcal{L}(\Xi_{A^*}^*)$ and $\text{PRM}(\lambda_{A^*}^*)$ for processes on the d -dimensional (normalised) space $E^* = (0, n]^d$.

Suppose that $\mathbf{U}, \mathbf{U}_1, \dots, \mathbf{U}_n$ are i.i.d. d -dimensional random vectors distributed according to an Archimedean copula $C(u_1, \dots, u_d) = \phi^{[-1]}(\phi(u_1) + \dots + \phi(u_d))$ on $E = [0, 1]^d$, satisfying (4.24), for some fixed parameter value $\theta > 1$. Let $\mathbf{V}, \mathbf{V}_1, \dots, \mathbf{V}_n$ be i.i.d. d -dimensional random vectors distributed according to a d -variate copula from family (2) (see Table 4.1) with parameter

θ , that we denote by C_0 ($:= C_{0,\theta}$):

$$\begin{aligned} C_0(u_1, \dots, u_d) &= \phi_0^{[-1]}(\phi_0(u_1) + \dots + \phi_0(u_d)) \\ &= \max \left\{ 1 - [(1 - u_1)^\theta + \dots + (1 - u_d)^\theta]^{\frac{1}{\theta}}, 0 \right\}. \end{aligned}$$

Suppose, for simplicity, that $s_{1n} = s_{2n} = \dots = s_{dn} = s_n \in (0, n]^d$, and define

$$A = A_n = \left[1 - \frac{s_n}{n}, 1 \right)^d \quad \text{and} \quad A^\star = A_n^\star = (0, s_n]^d.$$

Let $\Xi_{A^\star}^\star = \sum_{i=1}^n I_{\{\mathbf{U}_i^\star \in A^\star\}} \delta_{\mathbf{U}_i^\star}$ and $W_{A^\star}^\star = \sum_{i=1}^n I_{\{\mathbf{U}_i^\star \in A^\star\}}$, where the \mathbf{U}_i^\star 's are i.i.d. copies of $\mathbf{U}^\star = (n(1 - U_1), \dots, n(1 - U_d)) = n(\mathbf{1} - \mathbf{U})$. Furthermore, let $\mathbf{V}^\star = n(\mathbf{1} - \mathbf{V})$. As we chose i.i.d. \mathbf{U}_i 's, a bound on the error $d_{TV}(\mathcal{L}(\Xi_{A^\star}^\star), \text{PRM}(\mathbb{E}\Xi_{A^\star}^\star))$ is easy to determine: Proposition 4.2.3 gives the upper bound s_n/n . However, as for $d = 2$, we prefer to approximate further by $\text{PRM}(\boldsymbol{\lambda}_{A^\star}^\star)$, whose intensity function is given by

$$\lambda^\star(s_1, \dots, s_d) = \left(\prod_{j=1}^d [1 - (j-1)\theta] \right) (s_1 \dots s_d)^{\theta-1} (s_1^\theta + \dots + s_d^\theta)^{\frac{1}{\theta}-d},$$

for all $(s_1, \dots, s_d) \in E^\star = (0, n]^d$.

One possibility is to proceed analogously to the bivariate case treated in Theorem 4.3.4, where we determined an upper bound on $d_{TV}(\text{PRM}(\mathbb{E}\Xi_{A^\star}^\star), \text{PRM}(\boldsymbol{\lambda}_{A^\star}^\star))$ by straightforward comparison of their intensity functions, which led to rather involved computations. Computations would of course become even harder for $d \geq 3$ and we would need to introduce more and more assumptions on the functions w and h . For instance, w would need to be d -times continuously differentiable on $[0, 1)$.

Another possibility is to look for a cruder bound that does not require these assumptions. We can use the weaker d_2 -distance from Section 2.6. To achieve this note first that we can express A^\star as a union of d -rectangles of side lengths, say, s_n/m , for some integer $m \geq 1$. Define a d -rectangle as follows:

$$R_{\mathbf{k}}^\star := \left(k_1 \frac{s_n}{m}, (k_1 + 1) \frac{s_n}{m} \right] \times \dots \times \left(k_d \frac{s_n}{m}, (k_d + 1) \frac{s_n}{m} \right],$$

for all $\mathbf{k} = (k_1, \dots, k_d) \in A^\star$ with $k_j \in \{0, 1, \dots, m-1\}$, for each $1 \leq j \leq d$. The set A^\star is then a union of m^d d -rectangles. Then, define $P_{\mathbf{k}} = \mathcal{L}(\mathbf{U}^\star | \mathbf{U}^\star \in R_{\mathbf{k}}^\star)$

and $P'_{\mathbf{k}} = \mathcal{L}(\mathbf{V}^* | \mathbf{V}^* \in R_{\mathbf{k}}^*)$ for each \mathbf{k} , and note that by proceeding similarly as in the proof of Proposition 3.2.2, we can show that

$$\begin{aligned} \text{PRM}(\mathbb{E}\Xi_{R_{\mathbf{k}}^*}^*) &= \mathcal{L} \left(\sum_{j=1}^{N_{\mathbf{k}}} \delta_{\mathbf{Z}_{j,\mathbf{k}}} \right), \\ &\text{where } \mathbf{Z}_{j,\mathbf{k}} \stackrel{i.i.d.}{\sim} P_{\mathbf{k}}, \text{ independent of } N_{\mathbf{k}} \sim \text{Poi}(\mathbb{E}W_{R_{\mathbf{k}}^*}^*); \\ \text{PRM}(\boldsymbol{\lambda}_{R_{\mathbf{k}}^*}^*) &= \mathcal{L} \left(\sum_{j=1}^{L_{\mathbf{k}}} \delta_{\mathbf{Z}'_{j,\mathbf{k}}} \right), \\ &\text{where } \mathbf{Z}'_{j,\mathbf{k}} \stackrel{i.i.d.}{\sim} P'_{\mathbf{k}}, \text{ independent of } L_{\mathbf{k}} \sim \text{Poi}(\boldsymbol{\lambda}^*(R_{\mathbf{k}}^*)), \end{aligned}$$

where $\mathbb{E}W_{R_{\mathbf{k}}^*}^* = nP(\mathbf{U}^* \in R_{\mathbf{k}}^*)$ and $\boldsymbol{\lambda}^*(R_{\mathbf{k}}^*) = nP(\mathbf{V}^* \in R_{\mathbf{k}}^*)$. Now construct, for each \mathbf{k} , an additional Poisson process with intensity measure $\boldsymbol{\nu}_{R_{\mathbf{k}}^*}^*$ as follows:

$$\text{PRM}(\boldsymbol{\nu}_{R_{\mathbf{k}}^*}^*) = \mathcal{L} \left(\sum_{j=1}^{N_{\mathbf{k}}} \delta_{\mathbf{Z}'_{j,\mathbf{k}}} \right), \quad \text{where } \mathbf{Z}'_{j,\mathbf{k}} \stackrel{i.i.d.}{\sim} P'_{\mathbf{k}}, \text{ independent of } N_{\mathbf{k}}.$$

The corresponding Poisson processes on A^* can be realised in an analogous way, and, since Poisson processes constructed on disjoint sets are independent, we have that

$$\text{PRM}(\mathbb{E}\Xi_{A^*}^*) = \sum_{\mathbf{k}} \sum_{j=1}^{N_{\mathbf{k}}} \delta_{\mathbf{Z}_{j,\mathbf{k}}},$$

for all $\mathbf{k} = (k_1, \dots, k_d) \in A^*$ with $k_j \in \{0, 1, \dots, m-1\}$ for each $1 \leq j \leq d$, and similarly for $\text{PRM}(\boldsymbol{\lambda}_{A^*}^*)$ and $\text{PRM}(\boldsymbol{\nu}_{A^*}^*)$. The error that we want to estimate can then be split up into two parts:

$$\begin{aligned} d_2(\text{PRM}(\mathbb{E}\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\lambda}_{A^*}^*)) \\ \leq d_2(\text{PRM}(\mathbb{E}\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\nu}_{A^*}^*)) + d_2(\text{PRM}(\boldsymbol{\nu}_{A^*}^*), \text{PRM}(\boldsymbol{\lambda}_{A^*}^*)). \end{aligned}$$

Since $\text{PRM}(\mathbb{E}\Xi_{A^*}^*)$ and $\text{PRM}(\boldsymbol{\nu}_{A^*}^*)$ have the same number $nP(\mathbf{U}^* \in A^*)$ of expected points in A^* , Proposition 2.6.7 gives the upper bound $2d_1(\mathbb{E}\Xi_{A^*}^*, \boldsymbol{\nu}_{A^*}^*)$ for the first of the above two summands. With (2.38), the d_1 -distance may be bounded by the maximum d_0 -distance that points of the two Poisson processes may be apart, which is the length of the space diagonal of the d -rectangle in the

d_0 -distance. The d_0 -distance can be chosen in a way to give a good estimate. Choose, e.g., the Euclidean distance bounded by 1:

$$d_0(\mathbf{x}, \mathbf{y}) = \min \left(\sqrt{\sum_{j=1}^d (x_j - y_j)^2}, 1 \right),$$

for any $\mathbf{x} = (x_1, \dots, x_d)$ and $\mathbf{y} = (y_1, \dots, y_d) \in \mathbb{R}$. The d_0 -distance between two diagonally opposite corner points of $R_{\mathbf{k}}^*$ is then $\sqrt{d} s_n/m$ and we have the following estimate for the error caused by smearing out points over the rectangles:

$$d_2(\text{PRM}(\mathbb{E}\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\nu}_{A^*}^*)) \leq \frac{2\sqrt{d} s_n}{m}.$$

The second error term, i.e. $d_2(\text{PRM}(\boldsymbol{\nu}_{A^*}^*), \text{PRM}(\boldsymbol{\lambda}_{A^*}^*))$, may be estimated as follows:

$$\begin{aligned} & \sum_{\mathbf{k}} d_{TV}(\text{PRM}(\boldsymbol{\nu}_{R_{\mathbf{k}}^*}^*), \text{PRM}(\boldsymbol{\lambda}_{R_{\mathbf{k}}^*}^*)) \\ & \leq \sum_{\mathbf{k}} d_{TV}(\text{Poi}(nP(\mathbf{U}^* \in R_{\mathbf{k}}^*)), \text{Poi}(nP(\mathbf{V}^* \in R_{\mathbf{k}}^*))) \\ & \leq \sum_{\mathbf{k}} n|P(\mathbf{U}^* \in R_{\mathbf{k}}^*) - P(\mathbf{V}^* \in R_{\mathbf{k}}^*)|, \end{aligned} \quad (4.57)$$

where we use (2.42), as well as an argument similar to Michel's argument in the proof of Theorem 2.5.13 and the fact that $d_{TV}(\text{Poi}(\mu), \text{Poi}(\mu')) \leq |\mu - \mu'|$. Note that, with

$$P(\mathbf{U}^* \in R_{\mathbf{k}}^*) = \prod_{j=1}^d \mathbb{E}[I_{\{U_j^* \leq (k_j+1)s_n/m\}} - I_{\{U_j^* \leq k_j s_n/m\}}]$$

for each \mathbf{k} (and analogously for $P(\mathbf{V}^* \in R_{\mathbf{k}}^*)$), each of the two probabilities in (4.57) may be expressed as a sum of 2^d d -dimensional copulas, and

$$\begin{aligned} & |P(\mathbf{U}^* \in R_{\mathbf{k}}^*) - P(\mathbf{V}^* \in R_{\mathbf{k}}^*)| \\ & \leq 2^d \max_{(u_1, \dots, u_d) \in A} |C(u_1, \dots, u_d) - C_0(u_1, \dots, u_d)| \\ & = 2^d \max_{(s_1, \dots, s_d) \in A^*} \left| w^{-1} \left[\left(w^\theta \left(\frac{s_1}{n} \right) + \dots w^\theta \left(\frac{s_d}{n} \right) \right)^{\frac{1}{\theta}} \right] - \frac{1}{n} (s_1^\theta + \dots + s_d^\theta)^{\frac{1}{\theta}} \right|. \end{aligned} \quad (4.58)$$

As in Theorem 4.3.4, assume that $w(r) = rh(r)$ and that h has a positive derivative for all $r \in [0, \delta)$, for some $\delta > 0$ close to 0. We then have $h(0) \leq h(r) \leq h(0) + r \max_{0 \leq \xi \leq r} h'(\xi)$ for all $r \in [0, \delta)$. Also, for x small enough, we can show, with arguments similar to (4.32), that there is a constant $c > 0$ such that

$$\frac{x}{h(0)} - cx^2 \leq w^{-1}(x) \leq \frac{x}{h(0)}. \quad (4.59)$$

Supposing that s_n/n is small enough for (4.59) to be satisfied for $x := (w^\theta(s_1/n) + \dots + w^\theta(s_d/n))^{1/\theta}$, where $s_1, \dots, s_d \leq s_n$, it is then possible to show that

$$w^{-1} \left[\left(w^\theta \left(\frac{s_1}{n} \right) + \dots + w^\theta \left(\frac{s_d}{n} \right) \right)^{\frac{1}{\theta}} \right] = \frac{1}{n} (s_1^\theta + \dots + s_d^\theta)^{\frac{1}{\theta}} + O \left(\left(\frac{s_n}{n} \right)^2 \right).$$

Then there exists a constant $\alpha > 0$ such that (4.58) is smaller than $2^d \alpha (s_n/n)^2$, and we obtain the following estimate for the second error term:

$$d_2(\text{PRM}(\boldsymbol{\nu}_{A^*}^*), \text{PRM}(\boldsymbol{\lambda}_{A^*}^*)) \leq \frac{(2m)^d \alpha s_n^2}{n}.$$

The bound of the total error $d_2(\text{PRM}(\mathbb{E}\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\lambda}_{A^*}^*))$ is thus composed of a term of order s_n/m and another term of order $m^d s_n^2/n$. Now choose, for instance, $m = (n/s_n)^{1/(d+1)}$. Then both terms are of order $(s_n^{d+2}/n)^{1/(d+1)}$ and the total error is small only if $s_n \ll n^{1/(d+2)}$, i.e. if the threshold value s_n is smaller than a small power of n . This result is reminiscent of the result obtained in Theorem 4.3.4, which requires $s_n \ll n^{1/2}$.

4.4 MPPE's with bivariate Marshall-Olkin geometric marks

We consider MPPE's with bivariate marks that follow a certain bivariate geometric distribution, the Marshall-Olkin geometric distribution. We can readily approximate the law of this process by that of a Poisson process with the same mean measure by way of Theorem 4.0.1. However, as the marks have geometric, and thereby discrete margins, the mean measure will live on a lattice and be rather tedious to work with in practical applications. We would therefore prefer to approximate by a further Poisson process with a continuous mean measure. As the total variation distance is too strong for this kind of approximation, we use the weaker d_2 -distance instead, which is not as sensitive towards small

changes in the positions of the points of the point processes. As for MPPE's with univariate geometric marks, which we studied in Section 3.2.2, the error that arises when going from a process on a lattice to a process with continuous intensity will only be small if the parameters of the distribution of the marks vary with the sample size n at a suitable rate.

Section 4.4.1 introduces the bivariate Marshall-Olkin geometric distribution and relates it to its continuous counterpart, the bivariate Marshall-Olkin exponential distribution. Section 4.4.2 determines an error estimate in the total variation distance for the approximation of the law of the MPPE by that of a Poisson process with equal mean measure. In Section 4.4.3, we construct a continuous intensity function by spreading out the point probabilities of the Marshall-Olkin distribution over the entire space. As this intensity function depends on n , Section 4.4.4 makes some assumptions on the parameters of the Marshall-Olkin geometric distribution. These allow us to find another continuous intensity function that is asymptotically equal to the one that we constructed previously, but no longer varies with n . Section 4.4.5 establishes error estimates in the d_2 -distance for the approximation by the Poisson process whose intensity function we constructed in Section 4.4.3, whereas Section 4.4.6 gives error bounds, both in d_{TV} and in d_2 , for further approximating by a Poisson process with the intensity we found in Section 4.4.4. In Section 4.4.7, we summarise the results by adding up the d_2 -error bounds arising from each step, thus giving the total error bound for the approximation of the MPPE by the final Poisson process.

4.4.1 The bivariate Marshall-Olkin geometric distribution

The bivariate Marshall-Olkin geometric distribution arises as a natural generalisation of the geometric distribution to two dimensions. It was first introduced by Hawkes (1972) and later studied by Marshall and Olkin (1985) as the discrete counterpart to their bivariate exponential distribution, first derived by them in Marshall and Olkin (1967b) using shock models. Limit distributions for maxima of i.i.d. Marshall-Olkin geometric random pairs were established in Mitov and Nadarajah (2005) and Feidt et al. (2010).

Underlying the Marshall-Olkin geometric distribution are Bernoulli trials. Suppose S and T are two Bernoulli random variables with joint probability mass function $P(S = i, T = j) = p_{ij}$, for all $i, j = 0, 1$, and let S_1, S_2, \dots and T_1, T_2, \dots be i.i.d. copies of S and T , respectively. Let X_1 and X_2 denote the numbers of 0's before the first 1 in the sequences S_1, S_2, \dots and T_1, T_2, \dots ,

respectively. Obviously, X_1 and X_2 follow geometric distributions with failure probabilities $q_1 := P(S = 0) = p_{00} + p_{01}$ and $q_2 := P(T = 0) = p_{00} + p_{10}$, respectively. Their joint probability mass function is given by

$$P(X_1 = k, X_2 = l) = \begin{cases} p_{00}^k q_2^{l-k} (1 - p_{00}/q_2 - q_2 + p_{00}) & \text{for } k < l, \\ p_{00}^k (1 - q_1 - q_2 + p_{00}) & \text{for } k = l, \\ p_{00}^l q_1^{k-l} (1 - q_1 - p_{00}/q_1 + p_{00}) & \text{for } k > l, \end{cases} \quad (4.60)$$

for any $k, l \in \mathbb{Z}_+$. The distribution of $\mathbf{X} = (X_1, X_2)$ thus depends on three parameters: the two marginal failure probabilities q_1 and q_2 , as well as $p_{00} = P(S = 0, T = 0)$, the probability of joint failure. We assume that $p_{00} \geq q_1 q_2$. We have

$$P(X_1 \geq k, X_2 \geq l) = \begin{cases} p_{00}^k q_2^{l-k} & \text{for } k < l, \\ p_{00}^k & \text{for } k = l, \\ p_{00}^l q_1^{k-l} & \text{for } k > l. \end{cases} \quad (4.61)$$

The survival copula \hat{C} is given by a Marshall-Olkin copula $C_{\alpha, \beta}$ as defined in Example 4.1.1 (f). To show this, we may proceed as in Example 4.1.6 for the Marshall-Olkin exponential distribution. That is, rewrite (4.61) as

$$\left(\frac{p_{00}}{q_2}\right)^k \left(\frac{p_{00}}{q_1}\right)^l \left(\frac{q_1 q_2}{p_{00}}\right)^{\max(k, l)} = q_1^k q_2^l \min \left\{ \left(\frac{p_{00}}{q_1 q_2}\right)^k, \left(\frac{p_{00}}{q_1 q_2}\right)^l \right\},$$

using $\max(k, l) = k + l - \min(k, l)$ and $p_{00} \geq q_1 q_2$. With $u = P(X_1 \geq k) = q_1^k$, $v = P(X_2 \geq l) = q_2^l$, and

$$\alpha = \frac{\log(p_{00}/q_1 q_2)}{\log(1/q_1)}, \quad \beta = \frac{\log(p_{00}/q_1 q_2)}{\log(1/q_2)},$$

we have $(p_{00}/q_1 q_2)^k = u^{-\alpha}$ and $(p_{00}/q_1 q_2)^l = v^{-\beta}$, and $\hat{C}(u, v) = C_{\alpha, \beta}(u, v)$, for all $(u, v) \in (0, 1)^2$, with parameters $\alpha, \beta \in [0, 1]$ since $p_{00} \geq q_1 q_2$ and $q_1, q_2 \geq p_{00}$. For $\alpha, \beta \in (0, 1)$, the copulas in this family have full support, i.e. $[0, 1]^2$. Note that if $p_{00} = q_1 q_2$, the Marshall-Olkin geometric distribution corresponds to a bivariate distribution with independent geometric margins.

We can relate the Marshall-Olkin geometric distribution to its continuous counterpart, the Marshall-Olkin exponential distribution, by noting that (4.61) is equal to $P(\tilde{X}_1 \geq k, \tilde{X}_2 \geq l)$ for all $(k, l) \in \mathbb{Z}_+^2$ and for $(\tilde{X}_1, \tilde{X}_2)$ distributed according to (4.6) with parameters $\nu_1, \nu_2, \nu_{12} > 0$, $\nu := \nu_1 + \nu_2 + \nu_{12}$, if we set

$$q_1 := e^{-(\nu_1 + \nu_{12})}, \quad q_2 := e^{-(\nu_2 + \nu_{12})}, \quad p_{00} := e^{-\nu}.$$

The condition $p_{00} \geq q_1 q_2$ corresponds to $\nu_{12} \geq 0$. For $\nu_{12} = 0$, \tilde{X}_1 and \tilde{X}_2 are independent.

4.4.2 Approximation in d_{TV} by a Poisson process on a lattice

For any integer $n \geq 1$, let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be i.i.d. copies of the random pair $\mathbf{X} = (X_1, X_2)$, which follows the Marshall-Olkin geometric distribution from Section 4.4.1 and takes values in $\mathbb{Z}_+^2 \subset [0, \infty)^2$. Let $A \in \mathcal{B}([0, \infty)^2)$. We consider the MPPE $\Xi_A = \sum_{i=1}^n I_{\{\mathbf{X}_i \in A\}} \delta_{\mathbf{X}_i}$, which lives on the lattice \mathbb{Z}_+^2 . The following normalisation is the Marshall-Olkin geometric counterpart to the normalisation used in Section 4.2.3 for studying joint threshold exceedances of Marshall-Olkin exponential marks:

$$(k^*, l^*) = (k \log(1/p_{00}) - \log n, l \log(1/p_{00}) - \log n), \quad \text{for any } (k, l) \in \mathbb{Z}_+. \quad (4.62)$$

Under this normalisation, Ξ_A corresponds to

$$\Xi_{A^*}^* = \sum_{i=1}^n I_{\{\mathbf{X}_i^* \in A^*\}} \delta_{\mathbf{X}_i^*}, \quad (4.63)$$

which lives on the lattice E^* of normalised points (k^*, l^*) . Note that $E^* \subset [-\log n, \infty)^2$. Furthermore, denote by

$$W_{A^*}^* = \sum_{i=1}^n I_{\{\mathbf{X}_i^* \in A^*\}} \quad (4.64)$$

the random number of normalised points in A^* . For the particular choice $A = A_n = [u_n, \infty)^2$ for some threshold $u_n \in [0, \infty)$, we obtain $A^* = A_n^* = [u_n^*, \infty)^2$ with $u_n^* = u_n \log(1/p_{00}) - \log n$, and $\Xi_{A^*}^*$ captures joint threshold exceedances of the components of the normalised random pairs $\mathbf{X}_1^*, \dots, \mathbf{X}_n^*$.

The following proposition gives straightforward error estimates for the approximation of the law of $\Xi_{A^*}^*$ by that of a Poisson process with mean measure $\mathbb{E}\Xi_{A^*}^*$, both for general sets A^* , and for the particular choice $A^* = [u_n^*, \infty)^2$.

Proposition 4.4.1. *Suppose $\mathbf{X} = (X_1, X_2)$ follows the Marshall-Olkin geometric distribution with parameters $q_1, q_2, p_{00} \in (0, 1)$. For each integer $n \geq 1$, let $\mathbf{X}_1^*, \dots, \mathbf{X}_n^*$ be i.i.d. copies of the normalised random pair $\mathbf{X}^* = (X_1^*, X_2^*)$ with state space E^* , where $X_j^* = \log(1/p_{00})X_j - \log n$, for $j = 1, 2$. Let*

$A^* \in \mathcal{B}([0, \infty)^2)$ and let $\Xi_{A^*}^*$ and $W_{A^*}^*$ be defined as in (4.63) and (4.64), respectively. Then the mean measure of $\Xi_{A^*}^*$ is given by

$$\pi^*(B^*) := \pi_{A^*}^*(B^*) := \mathbb{E}\Xi_{A^*}^*(B^*) = \sum_{(k^*, l^*) \in A^* \cap E^* \cap B^*} nP(X_1^* = k^*, X_2^* = l^*),$$

for any $B^* \in \mathcal{B}([-\log n, \infty)^2)$, where, for any $(k^*, l^*) \in E^*$,

$$\begin{aligned} P(X_1^* = k^*, X_2^* = l^*) \\ = \begin{cases} \frac{1}{n} (1 - \frac{p_{00}}{q_2} - q_2 + p_{00}) e^{-\frac{\log(p_{00}/q_2)}{\log p_{00}} k^*} e^{-\frac{\log q_2}{\log p_{00}} l^*} & \text{for } k^* < l^*, \\ \frac{1}{n} (1 - q_1 - q_2 + p_{00}) e^{-k^*} & \text{for } k^* = l^*, \\ \frac{1}{n} (1 - q_1 - \frac{p_{00}}{q_1} + p_{00}) e^{-\frac{\log q_1}{\log p_{00}} k^*} e^{-\frac{\log(p_{00}/q_1)}{\log p_{00}} l^*} & \text{for } k^* > l^*, \end{cases} \quad (4.65) \end{aligned}$$

and $d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\pi^*)) \leq P(\mathbf{X}^* \in A^*)$. With $A^* = A_n^* = [u_n^*, \infty)^2$ for any choice of $u_n^* \geq -\log n$, we obtain

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\pi^*)) \leq \frac{e^{-u_n^*}}{n}. \quad (4.66)$$

Proof. With (4.60) and

$$P(X_1^* = k^*, X_2^* = l^*) = P\left(X_1 = \frac{k^* + \log n}{\log(1/p_{00})}, X_2 = \frac{l^* + \log n}{\log(1/p_{00})}\right),$$

we obtain (4.65) for the joint probability mass function of \mathbf{X}^* . For any set $B^* \in \mathcal{B}([-\log n, \infty)^2)$, the mean measure of $\Xi_{A^*}^*$ applied to B^* is then given by

$$nP(\mathbf{X}^* \in A^* \cap B^*) = \sum_{(k^*, l^*) \in A^* \cap E^* \cap B^*} nP(X_1^* = k^*, X_2^* = l^*).$$

By Theorem 4.0.1,

$$d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\pi^*)) \leq P(\mathbf{X}^* \in A^*),$$

where, using (3.36) and (4.61), we find

$$\begin{aligned}
 P(\mathbf{X}^* \in A^*) &= P(X_1^* \geq u_n^*, X_2^* \geq u_n^*) \\
 &= P\left(X_1 \geq \frac{u_n^* + \log n}{\log(1/p_{00})}, X_2 \geq \frac{u_n^* + \log n}{\log(1/p_{00}) l^*}\right) \\
 &= P\left(X_1 \geq \left\lceil \frac{u_n^* + \log n}{\log(1/p_{00})} \right\rceil, X_2 \geq \left\lceil \frac{u_n^* + \log n}{\log(1/p_{00})} \right\rceil\right) \\
 &= p_{00}^{\left\lceil \frac{u_n^* + \log n}{\log(1/p_{00})} \right\rceil} \leq \frac{e^{-u_n^*}}{n}.
 \end{aligned}$$

□

Remark 4.4.2. For $A^* = [u_{1n}^*, \infty) \times [u_{2n}^*, \infty)$ with $u_{1n}^* \neq u_{2n}^* \in [-\log n, \infty)$, we can proceed as in Section 4.2.3 for the Marshall-Olkin exponential distribution, i.e. use Theorem 4.0.1, in order to determine an estimate for $d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\pi^*))$.

The error bound in (4.66) is exactly the same as the one that we found in (4.14) for analogous MPPE's with Marshall-Olkin exponential instead of geometric marks. The difference is of course that the mean measure π^* of the MPPE with Marshall-Olkin geometric marks lives only on points $(k^*, l^*) \in A^* \cap E^*$ instead of on the whole of $A^* \cap [-\log n, \infty)^2$.

4.4.3 Construction of a “continuous” intensity function

Proposition 4.4.1 gives an error bound for the approximation of the MPPE $\Xi_{A^*}^*$ by a Poisson process whose mean measure $\mathbb{E}\Xi_{A^*}^*$ lives on the lattice of normalised points (k^*, l^*) , i.e. on

$$\begin{aligned}
 E^* &= \{(k^*, l^*) : k^* = k \log(1/p_{00}) - \log n, l^* = l \log(1/p_{00}) - \log n, \\
 &\quad \text{for all } (k, l) \in \mathbb{Z}_+^2\} \\
 &= (\log(1/p_{00})\mathbb{Z}_+ - \log n)^2 \subset [-\log n, \infty)^2.
 \end{aligned}$$

We would however prefer to approximate the law of the MPPE by that of a Poisson process with an easier-to-use and more flexible *continuous* intensity measure $\lambda^* = \lambda_{A^*}^*$ living on $A^* \cap [-\log n, \infty)^2$.

As discussed in Section 4.4.1, the survival copula of the Marshall-Olkin geometric distribution is a Marshall-Olkin copula, and thereby consists of both an

absolutely continuous part and a singular part on the curve $u^\alpha = v^\beta$ (which corresponds to the diagonal in $[-\log n, \infty)^2$). The “continuous” intensity measure λ^\star will have to mirror this behaviour, i.e. it will have to be of the form

$$\lambda^\star(B^\star) = \int_{A^\star \cap B^\star} \lambda^\star(s, t) ds dt + \int_{A^\star \cap B^\star \cap \{(s, t): s=t\}} \dot{\lambda}^\star(s) ds, \quad (4.67)$$

for any $B^\star \in \mathcal{B}([-\log n, \infty)^2)$, for “continuous” intensity functions λ^\star and $\dot{\lambda}^\star$ that, if integrated over the entire space, will give n , i.e. that will ensure that

$$\int_{-\log n}^{\infty} \int_{-\log n}^{\infty} \lambda^\star(s, t) ds dt + \int_{-\log n}^{\infty} \dot{\lambda}^\star(s) ds = nP(\mathbf{X}^\star \in [-\log n, \infty)^2) = n.$$

Remark 4.4.3. Note that for simplicity of language we here (and later on) somewhat abuse terminology when speaking of a “continuous” intensity function λ^\star or a “continuous” intensity measure λ^\star . The bivariate intensity function λ^\star is not continuous, but piecewise continuous, having a jump along the diagonal. The measure λ^\star is continuous only in the sense that it has an intensity with respect to Lebesgue measure (2-dimensional on the off-diagonal and 1-dimensional on the diagonal) and not with respect to a point measure.

The idea is to spread the point mass sitting on each of the off-diagonal lattice points $(k^\star, l^\star) \in E^\star$, $k^\star \neq l^\star$, uniformly over each of their corresponding coordinate rectangles (or rather, coordinate squares)

$$R_{k^\star, l^\star}^\star = \left[k^\star, k^\star + \log\left(\frac{1}{p_{00}}\right) \right) \times \left[l^\star, l^\star + \log\left(\frac{1}{p_{00}}\right) \right), \quad k^\star \neq l^\star,$$

and to also spread the point probabilities of the diagonal points (k^\star, k^\star) over the diagonal line $s = t$, where $s, t \geq -\log n$. We achieve this in the following three steps.

Step 1. Consider only the off-diagonal lattice points. We of course have

$$P(\mathbf{X}^\star \in R_{k^\star, l^\star}^\star) = P(X_1^\star = k^\star, X_2^\star = l^\star),$$

which is given by (4.65), and we may express the mean $nP(\mathbf{X}^\star \in A^\star)$ as

$$\begin{aligned} \sum_{(k^\star, l^\star) \in A^\star, k^\star \neq l^\star} n \int \int_{R_{k^\star, l^\star}^\star} \frac{P(X_1^\star = k^\star, X_2^\star = l^\star)}{\log^2(1/p_{00})} ds dt \\ + \sum_{(k^\star, k^\star) \in A^\star} nP(X_1^\star = k^\star, X_2^\star = k^\star), \end{aligned} \quad (4.68)$$

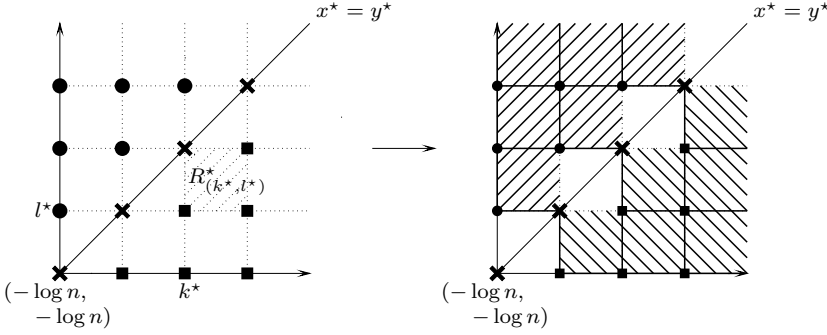


Figure 4.9: Spread the point masses on the off-diagonal points over the corresponding coordinate rectangles.

where $\log^2(1/p_{00})$ is the surface area of $R^*_{k^*, l^*}$. We have not actually changed anything yet as the integrand $P(X_1^* = k^*, X_2^* = l^*)/\log^2(1/p_{00})$ is constant with respect to the integrating variables s and t , and

$$\int \int_{R^*_{k^*, l^*}} \frac{P(X_1^* = k^*, X_2^* = l^*)}{\log^2(1/p_{00})} ds dt = P(X_1^* = k^*, X_2^* = l^*).$$

As we aim to find a continuous intensity function over the entire space $[-\log n, \infty)^2$, we exchange k^* and l^* in the expression of the point probability $P(X_1^* = k^*, X_2^* = l^*)$ from (4.65) by s and t , respectively. E.g., suppose that $k^* < l^*$. Then we replace the integral in (4.68) by

$$\int \int_{R^*_{k^*, l^*}} \frac{1 - p_{00}/q_2 - q_2 + p_{00}}{\log^2(1/p_{00})} e^{-\frac{\log(p_{00}/q_2)}{\log p_{00}} s} e^{-\frac{\log q_2}{\log p_{00}} t} ds dt \quad (4.69)$$

Evaluation of this new integral gives

$$\frac{1 - p_{00}/q_2 - q_2 + p_{00}}{\log(p_{00}/q_2) \log q_2} P(X_1^* = k^*, X_2^* = l^*). \quad (4.70)$$

The switch to variable s and t thus results only in the multiplication of the original point probability by a factor. The goal, however, is to integrate a

function in s and t over R_{k^*, l^*}^* and obtain the original point probability. This may be achieved by simply dividing the integrand in (4.69) by the multiplying factor found in (4.70). Hence, we rewrite the mean as follows

$$\sum_{(k^*, l^*) \in A^*, k^* \neq l^*} \int \int_{R_{k^*, l^*}^*} \lambda^*(s, t) ds dt + n \sum_{(k^*, k^*) \in A^*} P(X_1^* = k^*, X_2^* = k^*),$$

where

$$\lambda^*(s, t) = \frac{\log(p_{00}/q_2) \log q_2}{\log^2(1/p_{00})} e^{-\frac{\log(p_{00}/q_2)}{\log p_{00}} s} e^{-\frac{\log q_2}{\log p_{00}} t}, \forall (s, t) \in R_{k^*, l^*}^* \text{ with } k^* < l^*. \quad (4.71)$$

Analogously, we find

$$\lambda^*(s, t) = \frac{\log(p_{00}/q_1) \log q_1}{\log^2(1/p_{00})} e^{-\frac{\log q_1}{\log p_{00}} s} e^{-\frac{\log(p_{00}/q_1)}{\log p_{00}} t}, \forall (s, t) \in R_{k^*, l^*}^* \text{ with } k^* > l^*. \quad (4.72)$$

(4.71) and (4.72) supply suitable choices for the intensity function on coordinate rectangles lying above and below the diagonal, respectively. Figure 4.9 illustrates Step 1.

Step 2. We expand $\lambda^*(s, t)$ from (4.71) and (4.72) to the entire space (without the diagonal), i.e. we define

$$\lambda^*(s, t) := \begin{cases} \frac{\log(p_{00}/q_2) \log q_2}{\log^2(1/p_{00})} e^{-\frac{\log(p_{00}/q_2)}{\log p_{00}} s} e^{-\frac{\log q_2}{\log p_{00}} t} & \text{for } s < t, \\ \frac{\log(p_{00}/q_1) \log q_1}{\log^2(1/p_{00})} e^{-\frac{\log q_1}{\log p_{00}} s} e^{-\frac{\log(p_{00}/q_1)}{\log p_{00}} t} & \text{for } s > t, \end{cases}$$

for all $(s, t) \in [-\log n, \infty)^2$; see Figure 4.10. However, this adds surplus mass on the diagonal rectangles R_{k^*, k^*}^* .

Step 3. We adjust for the surplus mass on the diagonal rectangles by subtracting it from the point probabilities of the diagonal lattice points (k^*, k^*) , and accordingly rewrite the mean as follows:

$$\int \int_{A^*} \lambda_n^*(s, t) ds dt + n \sum_{(k^*, k^*) \in A^*} \left\{ P(X_1^* = k^*, X_2^* = k^*) - \frac{1}{n} \int \int_{R_{k^*, k^*}^*} \lambda_n^*(s, t) ds dt \right\}. \quad (4.73)$$

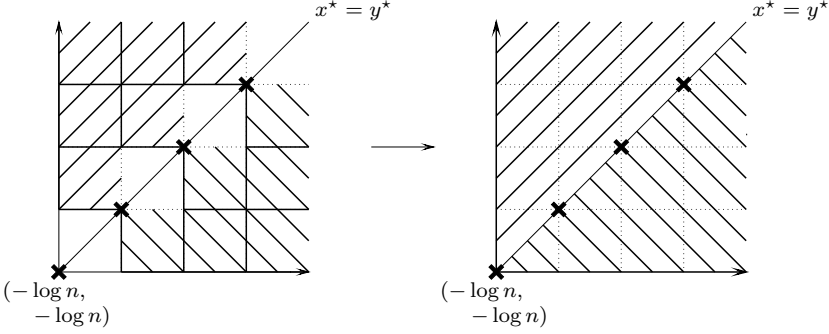


Figure 4.10: Define the intensity functions determined for off-diagonal rectangles on the entire space.

Computation of the term in curly brackets shows that the new mass that we put on the diagonal segments of each diagonal rectangle R_{k^*, k^*}^* is given by

$$\frac{e^{-k^*}}{n} (1 - p_{00}) \left[\frac{\log(1/q_1 q_2)}{\log(1/p_{00})} - 1 \right]. \quad (4.74)$$

Note that this equals

$$\int_{k^*}^{k^* + \log(1/p_{00})} \frac{e^{-s}}{n} \left[\frac{\log(1/q_1 q_2)}{\log(1/p_{00})} - 1 \right] ds,$$

for each $k^* \in E^*$, where we have parameterised the intensity function on the diagonal as projection along the s -axis. We thus define:

$$\lambda^*(s, t) = \begin{cases} \frac{\log(p_{00}/q_2) \log q_2}{\log^2(1/p_{00})} e^{-\frac{\log(p_{00}/q_2)}{\log p_{00}} s} e^{-\frac{\log q_2}{\log p_{00}} t} & \text{for } s < t, \\ \frac{\log(p_{00}/q_1) \log q_1}{\log^2(1/p_{00})} e^{-\frac{\log q_1}{\log p_{00}} s} e^{-\frac{\log(p_{00}/q_1)}{\log p_{00}} t} & \text{for } s > t, \end{cases} \quad (4.75)$$

$$\dot{\lambda}^*(s) = \frac{\log(p_{00}/q_1 q_2)}{\log(1/p_{00})} e^{-s} \quad \text{for } s = t.$$

Figure 4.11 illustrates this last step in the construction of λ^* .

The above construction guarantees the following:

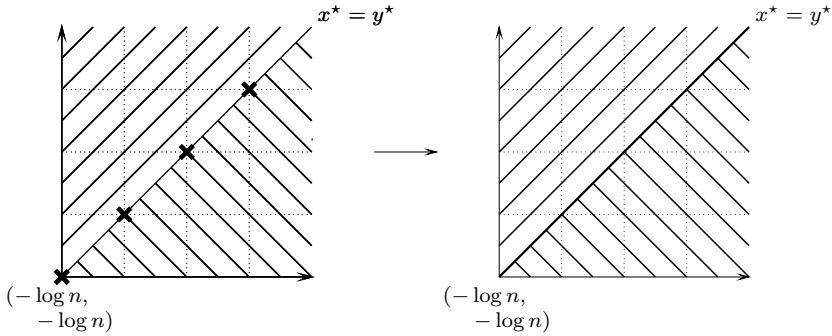


Figure 4.11: The mass on the diagonal lattice points is spread over the entire diagonal.

Proposition 4.4.4. Let λ^* , λ^* and $\dot{\lambda}^*$ be defined by (4.67) and (4.75). Then,

(i) $\lambda^*(R_{k^*, l^*}^*) = \pi^*(R_{k^*, l^*}^*)$, for any $(k^*, l^*) \in E^*$,

(ii) $\int_{[-\log n, \infty)^2} \lambda^*(s, t) ds dt + \int_{-\log n}^{\infty} \dot{\lambda}^*(s) ds = n$.

□

Remark 4.4.5. Proceeding as in Section 4.2.3 for the Marshall-Olkin exponential distribution, we may express the new intensity functions λ^* and $\dot{\lambda}^*$ in the original coordinate system by

$$\lambda(x, y) = \lambda_n(x, y) = \begin{cases} n \log(q_2) \log\left(\frac{p_{00}}{q_2}\right) p_{00}^x q_2^{y-x} & \text{for } x < y, \\ n \log(q_1) \log\left(\frac{p_{00}}{q_1}\right) q_1^{x-y} p_{00}^y & \text{for } x > y, \end{cases}$$

$$\dot{\lambda}(x) = \dot{\lambda}_n(x) = n \log\left(\frac{p_{00}}{q_1 q_2}\right) p_{00}^x \quad \text{for } x = y,$$

for any $(x, y) \in [0, \infty)^2$. We recognise a weighted and continuous version of $P(X_1 \geq k, X_2 \geq l)$ from (4.61).

4.4.4 Assumptions on the distributional parameters

The continuous intensity measure λ^* defined by (4.67) and (4.75) depends on the parameters q_1 , q_2 and p_{00} of the Marshall-Olkin geometric distribution. Our

aim is to determine a bound on the error for the approximation of the Poisson process with mean measure $\mathbb{E}\Xi_{A^*}^*$, living on the lattice E^* , by a Poisson process with mean measure λ^* . We already did something similar for the univariate geometric distribution in Proposition 3.2.9, and it turned out that the error could only become small if the success probability $p = p_n$ vanished as $n \rightarrow \infty$. As Section 4.4.5 will show below, the probability of simultaneous success, p_{11} , for the Marshall-Olkin geometric distribution, will similarly have to tend to 0 as $n \rightarrow \infty$. Since $p_{00} + p_{01} + p_{10} + p_{11} = 1$, this of course influences the distributional parameters p_{00} , q_1 and q_2 in that it also makes them dependent on n . The continuous intensity functions λ^* and $\dot{\lambda}^*$ thus have the drawback that, through their dependence on the parameters p_{00} , q_1 and q_2 , they are also dependent on n . We thus try to find other suitable continuous intensity functions that no longer vary with the sample size.

For simplicity, we make the assumption that p_{10} and p_{01} vary at the same rate as $p_{11} = p_{11n}$; more precisely, assume $p_{10} = p_{10n} = \gamma p_{11n}$ and $p_{01} = p_{01n} = \delta p_{11n}$, where γ and δ are strictly positive real numbers, bounded such that p_{10} and p_{01} are smaller than 1. We assume that p_{11n} tends to 0 as $n \rightarrow \infty$ at a rate that will be determined later, and express the distributional parameters as functions of it:

$$\begin{aligned} q_{1n} &= 1 - (1 + \gamma)p_{11n}, \\ q_{2n} &= 1 - (1 + \delta)p_{11n}, \\ p_{00n} &= 1 - (1 + \gamma + \delta)p_{11n}. \end{aligned} \tag{4.76}$$

Plugging into (4.75) and using the relation $\log(1-z) \sim -z$ for $|z| < 1$ and $z \rightarrow 0$, we find that $\lambda^*(s, t)$ and $\dot{\lambda}^*(s)$ are, for $p_{11n} \rightarrow 0$ as $n \rightarrow \infty$, asymptotically equal to

$$\begin{aligned} \lambda_{\gamma, \delta}^*(s, t) &:= \begin{cases} \frac{\gamma(1+\delta)}{(1+\gamma+\delta)^2} e^{-\frac{\gamma}{1+\gamma+\delta}s} e^{-\frac{1+\delta}{1+\gamma+\delta}t} & \text{for } s < t, \\ \frac{\delta(1+\gamma)}{(1+\gamma+\delta)^2} e^{-\frac{1+\gamma}{1+\gamma+\delta}s} e^{-\frac{\delta}{1+\gamma+\delta}t} & \text{for } s > t, \end{cases} \\ \text{and } \dot{\lambda}_{\gamma, \delta}^*(s) &:= \frac{1}{1 + \gamma + \delta} e^{-s} \quad \text{for } s = t, \end{aligned} \tag{4.77}$$

respectively, for all $(s, t) \in [-\log n, \infty)^2$. At first glance $\lambda_{\gamma, \delta}^*$ and $\dot{\lambda}_{\gamma, \delta}^*$ seem to be valid choices for continuous intensity functions independent of n . We will investigate in Section 4.4.6 whether a Poisson process with mean measure λ^*

on A^* may indeed be approximated by a Poisson process with mean measure

$$\lambda_{\gamma,\delta}^*(B^*) := \int \int_{A^* \cap B^*} \lambda_{\gamma,\delta}^*(s, t) ds dt + \int_{A^* \cap B^* \cap \{(s,t): s=t\}} \dot{\lambda}_{\gamma,\delta}^*(s) ds, \quad (4.78)$$

for all $B^* \in \mathcal{B}([-\log n, \infty)^2)$. To do the corresponding error calculations for a fixed sample size n we first need to examine in further detail the differences between the exponent terms in $\lambda^*(s, t)$ and $\lambda_{\gamma,\delta}^*(s, t)$:

Lemma 4.4.6. *For each integer $n \geq 1$, let $p_{11n} \in (0, 1)$ and let q_{1n} , q_{2n} , $p_{00n} \in (0, 1)$ be defined by (4.76). Then,*

$$\begin{aligned} (i) \quad & 0 \leq \frac{1 + \delta}{1 + \gamma + \delta} - \frac{\log q_{2n}}{\log p_{00n}} \leq \frac{\gamma p_{11n}}{1 - (1 + \gamma + \delta)p_{11n}}, \\ (ii) \quad & 0 \leq \frac{1 + \gamma}{1 + \gamma + \delta} - \frac{\log q_{1n}}{\log p_{00n}} \leq \frac{\delta p_{11n}}{1 - (1 + \gamma + \delta)p_{11n}}. \end{aligned}$$

Moreover,

$$\begin{aligned} (iii) \quad & 0 \leq \log \left(\frac{q_{2n}}{p_{00n}} \right) \leq \frac{\gamma p_{11n}}{1 - (1 + \gamma + \delta)p_{11n}}, \\ (iv) \quad & 0 \leq \log \left(\frac{q_{1n}}{p_{00n}} \right) \leq \frac{\delta p_{11n}}{1 - (1 + \gamma + \delta)p_{11n}}, \end{aligned}$$

and

$$(v) \quad \log \left(\frac{1}{p_{00n}} \right) \log \left(\frac{p_{00n}}{q_{1n}q_{2n}} \right) \leq \frac{(1 + \gamma + \delta)p_{11n}}{\{1 - (1 + \gamma + \delta)p_{11n}\}^2}.$$

Proof. (i) For ease of notation we omit the subscript n . Since, for all $|z| < 1$, $-\log(1 - z)/z$ is increasing and $-(1 - z)\log(1 - z)/z$ is decreasing, we obtain the following lower and upper bound, respectively, for $-(\log q_2)/(\log p_{00})$, where $q_2 < p_{00}$:

$$-\frac{1 + \delta}{1 + \gamma + \delta} \leq -\frac{\log q_2}{\log p_{00}} \leq -\frac{(1 + \delta) \cdot [1 - (1 + \gamma + \delta)p_{11}]}{[1 - (1 + \delta)p_{11}] \cdot (1 + \gamma + \delta)}.$$

Therefore,

$$\begin{aligned} 0 \leq \frac{1 + \delta}{1 + \gamma + \delta} - \frac{\log q_2}{\log p_{00}} & \leq \frac{1 + \delta}{1 + \gamma + \delta} \left\{ 1 - \frac{1 - (1 + \gamma + \delta)p_{11}}{1 - (1 + \delta)p_{11}} \right\} \\ & = \frac{(1 + \delta)\gamma p_{11}}{(1 + \gamma + \delta)[1 - (1 + \delta)p_{11}]} \leq \frac{\gamma p_{11}}{1 - (1 + \gamma + \delta)p_{11}}. \end{aligned}$$

(iii) Moreover, since $q_2 = p_{00} + p_{10}$, we have $\log(q_2/p_{00}) \geq 0$. Using $\log(1+z) \leq z$ for positive z , we obtain

$$\log(q_2/p_{00}) = \log\left(\frac{p_{00} + p_{10}}{p_{00}}\right) \leq \frac{p_{10}}{p_{00}} = \frac{\gamma p_{11}}{1 - (1 + \gamma + \delta)p_{11}}.$$

(ii) and (iv) can be shown analogously to (i) and (iii), respectively.

(v) We have

$$\begin{aligned} & \log\left(\frac{1}{p_{00n}}\right) \log\left(\frac{p_{00n}}{q_{1n}q_{2n}}\right) \\ &= (-\log p_{00}) \{-\log(p_{00} + p_{01}) - \log(p_{00} + p_{10}) + \log p_{00}\} \\ &\leq (-\log p_{00}) \{-\log p_{00} - \log p_{00} + \log p_{00}\} \\ &= (-\log p_{00})^2 \leq \frac{(1 - p_{00})^2}{p_{00}^2} \leq \frac{1 - p_{00}}{p_{00}^2} \\ &= \frac{(1 + \gamma + \delta)p_{11}}{\{1 - (1 + \gamma + \delta)p_{11}\}^2}. \end{aligned}$$

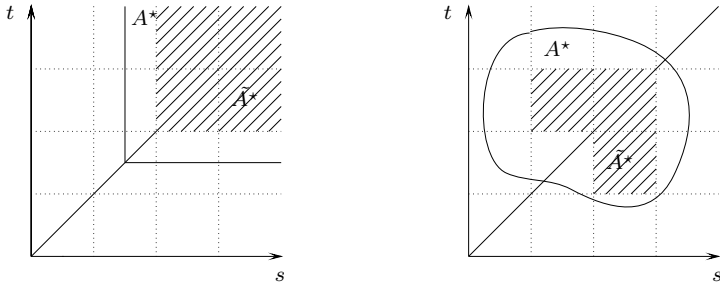
□

We will use Lemma 4.4.6 to determine error estimates in Sections 4.4.5 and 4.4.6.

Remark 4.4.7. We suppose here that γ and δ do not vary with n . However, the asymptotic equivalence of (4.75) and (4.77), and later results (i.e. Propositions 4.4.10 and 4.4.12, as well as Corollary 4.4.13) also hold for the case $\gamma = \gamma_n$ and $\delta = \delta_n$. These results are thus actually stronger than we make them out to be.

4.4.5 Approximation in d_2 by a Poisson process with continuous intensity

We now determine the error of the approximation of the Poisson process with mean measure π^* , living on lattice points $(k^*, l^*) \in A^* \cap E^*$, and the Poisson process with continuous mean measure λ^* , living on $A^* \cap [-\log n, \infty)^2$. As the total variation distance is too strong to achieve this, we use the weaker d_2 -distance that we introduced in Section 2.6. Theorem 4.4.8 gives a general error estimate for any set $A^* \in \mathcal{B}([-\log n, \infty)^2)$, which Proposition 4.4.10 in turn applies to the particular choice $A^* = [u^*, \infty)^2$.


 Figure 4.12: Examples of sets \tilde{A}^* .

Note that any not too small set $A^* \in \mathcal{B}([-\log n, \infty)^2)$ contains subsets that are unions of coordinate rectangles R_{k^*, l^*}^* , i.e. of the form

$$\bigcup_{(k^*, l^*) \in M^*} R_{k^*, l^*}^* \subseteq A^*, \quad (4.79)$$

where M^* is a countable subset of E^* . Let \tilde{A}^* denote the biggest set $\subseteq A^*$ of the form (4.79); see Figure 4.12 for some examples. In order to prove Theorem 4.4.8, we distinguish between the errors on \tilde{A}^* and $A^* \setminus \tilde{A}^*$. Even though $\pi^*(A^*) = \lambda^*(A^*)$ is not necessarily satisfied, Proposition 4.4.4 ensures that at least $\pi^*(\tilde{A}^*) = \lambda^*(\tilde{A}^*)$. We may therefore use Lemma 2.6.6 to bound the error on \tilde{A}^* by way of the d_1 -distance between π^* and λ^* on \tilde{A}^* . The size of the d_1 -distance depends on the choice of the d_0 -distance. As in Section 3.2.2, where we treated MPPE's with univariate geometric marks, we choose the Euclidean distance bounded by 1. For the remaining error, we rely on the “small” size of $A^* \setminus \tilde{A}^*$ and use Lemma 2.6.3 for an upper bound on $\Delta_1 \gamma$, where γ is the solution to an appropriate Stein equation.

Theorem 4.4.8. *With the notations from Sections 4.4.1-4.4.4, we obtain, for a set $A^* \in \mathcal{B}([-\log n, \infty)^2)$,*

$$d_2(\text{PRM}(\pi^*), \text{PRM}(\lambda^*)) \leq 2\sqrt{2} \log(1/p_{00}) + \left(1 \wedge \frac{1.65}{\sqrt{\lambda^*(A^*)}}\right) \lambda^*(A^* \setminus \tilde{A}^*), \quad (4.80)$$

where \tilde{A}^* denotes the biggest set $\subseteq A^*$ that is a union of coordinate rectangles, i.e. $\tilde{A}^* = \cup_{(k^*, l^*) \in M^*} R_{k^*, l^*}^*$, where M^* is the biggest subset of E^* such that $\tilde{A}^* \subseteq A^*$.

Proof. Let $\Xi_{\pi^*} \sim \text{PRM}(\pi^*)$ and $\Xi_{\lambda^*} \sim \text{PRM}(\lambda^*)$. Suppose that $Z = \{Z_t, t \in \mathbb{R}_+\}$ is an immigration-death process on A^* with immigration intensity λ^* , unit per-capita death rate, equilibrium distribution $\mathcal{L}(\Xi_{\lambda^*})$, and generator \mathcal{A} . Furthermore, let \mathcal{H} denote the set of functions $h : M_p(A^*) \rightarrow \mathbb{R}$ such that (2.40) is satisfied and let $\gamma : M_p(A^*) \rightarrow \mathbb{R}$ be defined by $\gamma(\xi) = -\int_0^\infty \{\mathbb{E}^\xi h(Z_t) - \text{PRM}(\lambda^*)\} dt$, for any $\xi \in M_p(A^*)$. By Proposition 2.5.4, γ is well-defined, and by (2.43), $|\text{PRM}(\pi^*)(h) - \text{PRM}(\lambda^*)(h)|$ equals $|\mathbb{E}(\mathcal{A}\gamma)(\Xi_{\pi^*})|$. Proceeding as in the proof of Proposition 2.5.10, we find that

$$\mathbb{E}(\mathcal{A}\gamma)(\Xi_{\pi^*}) = \mathbb{E} \int_{A^*} [\gamma(\Xi_{\pi^*} + \delta_{\mathbf{z}}) - \gamma(\Xi_{\pi^*})] (\lambda^*(d\mathbf{z}) - \pi^*(d\mathbf{z})),$$

and thus

$$\frac{|\mathbb{E}(\mathcal{A}\gamma)(\Xi_{\pi^*})|}{s_2(h)} \leq \frac{1}{s_2(h)} \mathbb{E} \left| \int_{A^*} [\gamma(\Xi_{\pi^*} + \delta_{\mathbf{z}}) - \gamma(\Xi_{\pi^*})] (\lambda^*(d\mathbf{z}) - \pi^*(d\mathbf{z})) \right|,$$

where, for any $\xi \in M_p(A^*)$,

$$\begin{aligned} & \left| \int_{A^*} [\gamma(\xi + \delta_{\mathbf{z}}) - \gamma(\xi)] (\lambda^*(d\mathbf{z}) - \pi^*(d\mathbf{z})) \right| \\ & \leq \left| \int_{\tilde{A}^*} [\gamma(\xi + \delta_{\mathbf{z}}) - \gamma(\xi)] (\lambda^*(d\mathbf{z}) - \pi^*(d\mathbf{z})) \right| \\ & \quad + \left| \int_{A^* \setminus \tilde{A}^*} [\gamma(\xi + \delta_{\mathbf{z}}) - \gamma(\xi)] (\lambda^*(d\mathbf{z}) - \pi^*(d\mathbf{z})) \right|. \end{aligned} \quad (4.81)$$

The second summand may be bounded by

$$\int_{A^* \setminus \tilde{A}^*} |\gamma(\xi + \delta_{\mathbf{z}}) - \gamma(\xi)| \cdot |\lambda^*(d\mathbf{z}) - \pi^*(d\mathbf{z})| \leq \Delta_1 \gamma \int_{A^* \setminus \tilde{A}^*} |\lambda^*(d\mathbf{z}) - \pi^*(d\mathbf{z})|. \quad (4.82)$$

Note that

$$\int_{A^* \setminus \tilde{A}^*} |\lambda^*(d\mathbf{z}) - \pi^*(d\mathbf{z})| \leq \lambda^*(A^* \setminus \tilde{A}^*),$$

and that Lemma 2.6.3 gives

$$\Delta_1 \gamma \leq s_2(h) \left(1 \wedge \frac{1.65}{\sqrt{\lambda^*(A^*)}} \right). \quad (4.83)$$

By Proposition 4.4.4, $\lambda^*(\tilde{A}^*) = \pi^*(\tilde{A}^*) = nP(\mathbf{X}^* \in \tilde{A}^*) < \infty$. We may therefore use Lemma 2.6.6 to bound the first summand by

$$s_2(h) \left(1 - e^{-\lambda^*(\tilde{A}^*)} \right) \left(1 + \frac{\lambda^*(\tilde{A}^*)}{|\xi| + 1} \right) d_1(\pi^*, \lambda^*)|_{\tilde{A}^*}, \quad (4.84)$$

where $d_1(\cdot, \cdot)|_{\tilde{A}^*}$ denotes the d_1 -distance on \tilde{A}^* (instead of on A^*). We have

$$\mathbb{E} \left(\frac{1}{|\Xi_{\pi^*}| + 1} \right) = \frac{1 - e^{-\pi^*(A^*)}}{\pi^*(A^*)}, \quad (4.85)$$

since $|\Xi_{\pi^*}| \sim \text{Poi}(\pi^*(A^*))$. Taking expectations in (4.81) and using (4.82) - (4.85), we obtain

$$\begin{aligned} & |\mathbb{E}(\mathcal{A}\gamma)(\Xi_{\pi^*})| / s_2(h) \\ & \leq \left(1 - e^{-\lambda^*(\tilde{A}^*)} \right) \left\{ 1 + \frac{\lambda^*(\tilde{A}^*)}{\pi^*(A^*)} \left(1 - e^{-\pi^*(A^*)} \right) \right\} d_1(\pi^*, \lambda^*)|_{\tilde{A}^*} \\ & \quad + \left(1 \wedge \frac{1.65}{\sqrt{\lambda^*(A^*)}} \right) \lambda^*(A^* \setminus \tilde{A}^*). \end{aligned} \quad (4.86)$$

We may further simplify by bounding $1 - e^{-\lambda^*(\tilde{A}^*)}$ and $1 - e^{-\pi^*(A^*)}$ by 1 and noting that, since $\pi^*(A^*) = \pi^*(A^* \setminus \tilde{A}^*) + \lambda^*(\tilde{A}^*) \geq \lambda^*(\tilde{A}^*)$, we have

$$1 + \frac{\lambda^*(\tilde{A}^*)}{\pi^*(A^*)} \leq 2. \quad (4.87)$$

With the definitions of \mathcal{K} and $s_1(\kappa)$ from Section 2.6, the d_1 -distance between λ^* and π^* on \tilde{A}^* is given by

$$d_1(\pi^*, \lambda^*)|_{\tilde{A}^*} = \frac{1}{\lambda^*(\tilde{A}^*)} \sup_{\kappa \in \mathcal{K}} \frac{\left| \int_{\tilde{A}^*} \kappa d\pi^* - \int_{\tilde{A}^*} \kappa d\lambda^* \right|}{s_1(\kappa)}.$$

As \tilde{A}^* is a union of coordinate rectangles R_{k^*, l^*}^* , the term $\int_{\tilde{A}^*} \kappa d\pi^* - \int_{\tilde{A}^*} \kappa d\lambda^*$ may be expressed as

$$\sum_{(k^*, l^*) \in \tilde{A}^*} \left\{ \int_{R_{k^*, l^*}^*} \kappa(\mathbf{z}) \pi^*(d\mathbf{z}) - \int_{R_{k^*, l^*}^*} \kappa(\mathbf{z}) \lambda^*(d\mathbf{z}) \right\}. \quad (4.88)$$

Furthermore, again by Proposition 4.4.4,

$$\int_{R_{k^*, l^*}^*} \kappa(\mathbf{z}) \pi^*(d\mathbf{z}) = \kappa((k^*, l^*)) \pi^*(R_{k^*, l^*}^*) = \kappa((k^*, l^*)) \lambda^*(R_{k^*, l^*}^*).$$

Hence, we find the following upper bound for (4.88):

$$\sum_{(k^*, l^*) \in \tilde{A}^*} \int_{R_{k^*, l^*}^*} |\kappa((k^*, l^*)) - \kappa(\mathbf{z})| \lambda^*(d\mathbf{z}),$$

which, by definition of the Lipschitz constant $s_1(k)$, is smaller than

$$s_1(\kappa) d_0((k^*, l^*), \mathbf{z}) \lambda^*(\tilde{A}^*).$$

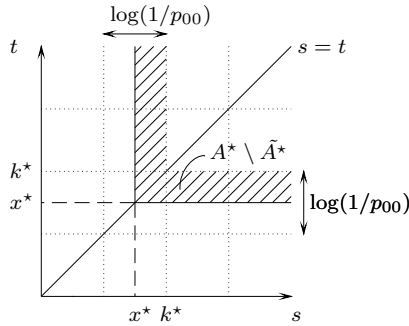
The biggest possible Euclidean distance between the lower left corner point (k^*, l^*) and any other point \mathbf{z} in the rectangle R_{k^*, l^*}^* is given by the length $\sqrt{2} \log(1/p_{00})$ of its diagonal. Thus,

$$d_1(\pi^*, \lambda^*)|_{\tilde{A}^*} \leq \sqrt{2} \log(1/p_{00}). \quad (4.89)$$

(4.87) and (4.89) give the upper bound $2\sqrt{2} \log(1/p_{00})$ for the first summand of the error term in (4.86). This completes the proof. \square

Theorem 4.4.8 gives sharp results only if the probability of simultaneous failure, $p_{00} = p_{00n}$ tends to 1 as $n \rightarrow \infty$. This makes sense since $\log(1/p_{00})$, introduced as scaling factor of the original marginal geometric random variables, provides the side lengths of the rescaled lattice squares. The condition $p_{00n} \uparrow 1$ makes the side lengths of the coordinate squares tend to 0 and thus causes the “disappearance” of the lattice into the whole real subset $[-\log n, \infty)^2$. The same holds for the area $A^* \setminus \tilde{A}^*$, thereby also causing the disappearance of the second error term as $n \rightarrow \infty$.

For sets A^* that are unions of coordinate rectangles, we immediately obtain the following corollary, as there is no left-over area $A^* \setminus \tilde{A}^*$, and by consequence no second error term.


 Figure 4.13: The set $A^* \setminus \tilde{A}^*$.

Corollary 4.4.9. *Let $A^* \in \mathcal{B}([-\log n, \infty)^2)$ be a union of coordinate rectangles, i.e. $A^* = \cup_{(k^*, l^*) \in M^*} R_{k^*, l^*}^*$ where $M^* \subseteq E^*$. Then,*

$$d_2(\text{PRM}(\boldsymbol{\pi}^*), \text{PRM}(\boldsymbol{\lambda}^*)) \leq 2\sqrt{2} \log(1/p_{00}).$$

□

We now apply Theorem 4.4.8 to the case where $A^* = A_n^* = [u_n^*, \infty)^2$ and express the error estimate in terms of the threshold u_n^* and the probability of simultaneous success p_{11n} . To achieve this we assume that the distributional parameters p_{00} , q_1 and q_2 are defined as in Section 4.4.4.

Proposition 4.4.10. *Let $p_{11n} \in (0, 1)$ and assume that q_{1n} , q_{2n} and p_{00n} satisfy (4.76). For any choice of $u_n^* \geq -\log n$, define $A^* = [u_n^*, \infty)^2$. With the notations from Theorem 4.4.8,*

$$\begin{aligned} d_2(\text{PRM}(\boldsymbol{\pi}^*), \text{PRM}(\boldsymbol{\lambda}^*)) \\ \leq \frac{(1 + \gamma + \delta)p_{11n}}{[1 - (1 + \gamma + \delta)p_{11n}]^2} \left\{ 2\sqrt{2} + 3 \left(e^{-u_n^*} \wedge 1.65e^{-u_n^*/2} \right) \right\}. \end{aligned}$$

Proof. For ease of notation we omit the subscript n . We apply result (4.80) from Theorem 4.4.8 to the special case $A^* = [u^*, \infty)^2$. Due to (4.76) and $-\log(1 - z) \leq z/(1 - z)$ for $|z| < 1$, we may bound the first of the two error

terms in (4.80) as follows:

$$2\sqrt{2}\log(1/p_{00}) \leq \frac{2\sqrt{2}(1+\gamma+\delta)p_{11}}{1-(1+\gamma+\delta)p_{11}} \leq \frac{2\sqrt{2}(1+\gamma+\delta)p_{11}}{[1-(1+\gamma+\delta)p_{11}]^2}. \quad (4.90)$$

Direct computation yields $\lambda^*(A^*) = e^{-u^*}$. As illustrated by Figure 4.13, $\lambda^*(A^* \setminus \tilde{A}^*)$ may be bounded by

$$\begin{aligned} & \int_{u^*}^{\infty} \log(1/p_{00}) \sup_{s \in [u^*, k^*]} \lambda^*(s, t) dt \\ & + \int_{u^*}^{\infty} \log(1/p_{00}) \sup_{t \in [u^*, k^*]} \lambda^*(s, t) ds + \int_{u^*}^{k^*} \sqrt{2}\log(1/p_{00}) \sup_{s \in [u^*, k^*]} \dot{\lambda}^*(s) ds. \end{aligned}$$

Note that

$$\begin{aligned} \sup_{s \in [u^*, k^*]} \exp \left\{ -\frac{\log(p_{00}/q_2)}{\log p_{00}} s \right\} & \leq \exp \left\{ -\frac{\log(p_{00}/q_2)}{\log p_{00}} u^* \right\}, \\ \sup_{t \in [u^*, k^*]} \exp \left\{ -\frac{\log(p_{00}/q_1)}{\log p_{00}} t \right\} & \leq \exp \left\{ -\frac{\log(p_{00}/q_1)}{\log p_{00}} u^* \right\}, \end{aligned}$$

and $\sup_{s \in [u^*, k^*]} e^{-s} \leq e^{-u^*}$. Thus, by definition (4.75) of $\lambda^*(s, t)$,

$$\begin{aligned} & \int_{u^*}^{\infty} \log(1/p_{00}) \sup_{s \in [u^*, k^*]} \lambda^*(s, t) dt \\ & \leq \frac{\log(p_{00}/q_2) \log q_2}{\log(1/p_{00})} e^{-\frac{\log(p_{00}/q_2)}{\log p_{00}} u^*} \int_{u^*}^{\infty} e^{-\frac{\log q_2}{\log p_{00}} t} dt, \end{aligned}$$

which equals $\log(q_2/p_{00})e^{-u^*}$. Analogously,

$$\int_{u^*}^{\infty} \log(1/p_{00}) \sup_{t \in [u^*, k^*]} \lambda^*(s, t) ds \leq \log(q_1/p_{00})e^{-u^*},$$

whereas

$$\int_{u^*}^{k^*} \sqrt{2}\log(1/p_{00}) \sup_{s \in [u^*, k^*]} \dot{\lambda}^*(s) ds \leq 2\log^2(1/p_{00}) \frac{\log(p_{00}/q_1 q_2)}{\log(1/p_{00})} e^{-u^*},$$

since $k^* - u^* \leq \sqrt{2} \log(1/p_{00})$. We obtain

$$\lambda^*(A^* \setminus \tilde{A}^*) = e^{-u^*} \left\{ \log \left(\frac{q_2}{p_{00}} \right) + \log \left(\frac{q_1}{p_{00}} \right) + 2 \log \left(\frac{1}{p_{00}} \right) \log \left(\frac{p_{00}}{q_1 q_2} \right) \right\}.$$

By Lemma 4.4.6 (iii)-(v), the term in curly brackets may be bounded by

$$\frac{(\gamma + \delta)p_{11}}{1 - (1 + \gamma + \delta)p_{11}} + \frac{2(1 + \gamma + \delta)p_{11}}{\{1 - (1 + \gamma + \delta)p_{11}\}^2} \leq \frac{3(1 + \gamma + \delta)p_{11}}{\{1 - (1 + \gamma + \delta)p_{11}\}^2}.$$

An upper bound for the second error term in (4.80) is thus given by

$$\left(e^{-u_n^*} \wedge 1.65e^{-u_n^*/2} \right) \frac{3(1 + \gamma + \delta)p_{11n}}{[1 - (1 + \gamma + \delta)p_{11n}]^2}.$$

By adding this to the bound in (4.90) we obtain the result. \square

The first of the error terms given by Proposition 4.4.10, i.e.

$$\frac{2\sqrt{2}(1 + \gamma + \delta)p_{11n}}{[1 - (1 + \gamma + \delta)p_{11n}]^2},$$

is a bound on the error $2\sqrt{2} \log(1/p_{00n})$ from Theorem 4.4.8, where we used the assumption from Section 4.4.4 that $p_{00n} = 1 - (1 + \gamma + \delta)p_{11n}$. This error term thus becomes small only if the probability of simultaneous success, p_{11n} , tends to 0 as n increases. The second error term, i.e.

$$\left(e^{-u_n^*} \wedge 1.65e^{-u_n^*/2} \right) \frac{3(1 + \gamma + \delta)p_{11n}}{[1 - (1 + \gamma + \delta)p_{11n}]^2},$$

is the bigger of the two, and determines the rate at which p_{11n} must converge to 0. The reason for that is that p_{11n} must converge fast enough in order to offset the effect of the factor $e^{-u_n^*}$ which we will want to be increasing with increasing n , since $e^{-u_n^*} = \lambda^*(A_n^*)$ is the expected number of points in A_n^* of the approximating Poisson process, as well as more or less the expected number of threshold exceedances of the MPPE, for which we have $e^{-u_n^*}/p_{00n} \leq \pi^*(A_n^*) \leq e^{-u_n^*}$. For instance, for a threshold u_n^* of size $-\log \log n$, the expected number of points in A^* of the two Poisson processes is $\log n$, the MPPE captures roughly the biggest $\log n$ points of its sample, and we need $p_{11n} = o(\log^{-1} n)$ for a sharp error bound. Suppose, for example, that $p_{11n} = n^{-1}$. Then, by

(4.76), the marginal probabilities of failure of \mathbf{X}_n^* , q_{1n} and q_{2n} , as well as the probability of simultaneous failure, p_{00n} , tend to 1 very fast.

The mean measure λ^* is by definition dependent on the values of the distributional parameters. Since these need to vary with the sample size n in order to obtain a small error for the approximation of $\text{PRM}(\pi^*)$ by $\text{PRM}(\lambda^*)$, it follows that $\lambda^* = \lambda_n^*$ (and of course also $\pi^* = \pi_n^*$). Though we have now achieved the goal of successfully approximating by a Poisson process with a continuous intensity, the conditions needed to accomplish this imply that we are not satisfied with our results yet, since we prefer to approximate by a Poisson process with continuous intensity that does not vary with n . As the next section will demonstrate, a suitable candidate is given by the Poisson process with intensity measure $\lambda_{\gamma,\delta}^*$ defined in (4.78).

4.4.6 Approximation in d_2 and d_{TV} by a Poisson process independent of n

We determine an error estimate for the approximation of the Poisson process with intensity measure $\lambda^* = \lambda_n^*$ by the Poisson process with intensity measure $\lambda_{\gamma,\delta}^*$, defined in (4.78), that does not depend on the sample size n . Since both intensities are continuous, there is no special need to use the d_2 -distance. We give the error in both the total variation and the d_2 distances. For the error in total variation we may straightforwardly use Proposition 2.5.10 for the approximation of two Poisson processes. For the d_2 -error, which will be smaller than the d_{TV} , we may additionally use Lemma 2.6.3 for an upper bound on $\Delta_1\gamma$, where γ is the solution of an adequate Stein equation. This bound, containing the factor $\lambda^*(A^*)^{-1/2}$ (or $\lambda_{\gamma,\delta}^*(A^*)^{-1/2}$), serves in reducing the d_2 -error. The error bounds given by Theorem 4.4.11 will become small for large n due to the pointwise convergence of the intensity functions $\lambda_n^*(s, t)$ and $\dot{\lambda}_n^*(s)$ to the intensity functions $\lambda_{\gamma,\delta}^*(s, t)$ and $\dot{\lambda}_{\gamma,\delta}^*(s)$, respectively, as $n \rightarrow \infty$.

Theorem 4.4.11. *With the notations from Sections 4.4.1-4.4.4, we obtain, for any set $A^* \in \mathcal{B}([-\log n, \infty)^2)$,*

$$\begin{aligned} (i) \quad d_{TV}(\text{PRM}(\lambda^*), \text{PRM}(\lambda_{\gamma,\delta}^*)) &\leq \int_{A^*} |\lambda^*(d\mathbf{z}) - \lambda_{\gamma,\delta}^*(d\mathbf{z})|, \\ (ii) \quad d_2(\text{PRM}(\lambda^*), \text{PRM}(\lambda_{\gamma,\delta}^*)) &\leq \left(1 \wedge 1.65 \min \left\{ \lambda^*(A^*)^{-1/2}, \lambda_{\gamma,\delta}^*(A^*)^{-1/2} \right\}\right) \int_{A^*} |\lambda^*(d\mathbf{z}) - \lambda_{\gamma,\delta}^*(d\mathbf{z})|. \end{aligned}$$

Proof. (i) By Proposition 4.4.4 (ii), λ^* is finite. Moreover, $\lambda_{\gamma,\delta}^*$ is finite since integration of $\lambda_{\gamma,\delta}^*$ and $\dot{\lambda}_{\gamma,\delta}^*$ over $[u^*, \infty)^2$ gives

$$\begin{aligned} & \int_{u^*}^{\infty} dt \int_{u^*}^t ds \frac{\gamma(1+\delta)}{(1+\gamma+\delta)^2} e^{-\frac{\gamma}{1+\gamma+\delta}s} e^{-\frac{1+\delta}{1+\gamma+\delta}t} \\ & + \int_{u^*}^{\infty} ds \int_{u^*}^s dt \frac{\delta(1+\gamma)}{(1+\gamma+\delta)^2} e^{-\frac{1+\gamma}{1+\gamma+\delta}s} e^{-\frac{\delta}{1+\gamma+\delta}t} + \int_{u^*}^{\infty} ds \frac{1}{1+\gamma+\delta} e^{-s} \\ & = \frac{\gamma}{1+\gamma+\delta} e^{-u^*} + \frac{\delta}{1+\gamma+\delta} e^{-u^*} + \frac{1}{1+\gamma+\delta} e^{-u^*} = e^{-u^*}, \end{aligned}$$

which equals n for $u^* = -\log n$. Proposition 2.5.10 then immediately gives the result.

(ii) Using the same immigration-death process Z and arguments as in the proof of Theorem 4.4.8, we can show that for $\Xi_{\gamma,\delta}^* \sim \text{PRM}(\lambda_{\gamma,\delta}^*)$,

$$\begin{aligned} \mathbb{E}h(\Xi_{\gamma,\delta}^*) - \text{PRM}(\lambda^*)(h) \\ = \mathbb{E} \left\{ \int_{A^*} [\gamma(\Xi_{\gamma,\delta}^* + \delta_{\mathbf{z}}) - \gamma(\Xi_{\gamma,\delta}^*)](\lambda^*(d\mathbf{z}) - \lambda_{\gamma,\delta}^*(d\mathbf{z})) \right\}. \end{aligned}$$

Analogously to (4.82) and (4.83), the integrand may be bounded by

$$\Delta_1 \gamma \int_{A^*} |\lambda^*(d\mathbf{z}) - \lambda_{\gamma,\delta}^*(d\mathbf{z})| \leq s_2(h) \left(1 \wedge \frac{1.65}{\sqrt{\lambda^*(A^*)}} \right) \int_{A^*} |\lambda^*(d\mathbf{z}) - \lambda_{\gamma,\delta}^*(d\mathbf{z})|.$$

Here, $1.65(\lambda^*(A^*))^{-\frac{1}{2}}$ may be replaced by $1.65(\lambda_{\gamma,\delta}^*(A^*))^{-\frac{1}{2}}$ by going through the same arguments as before, but instead starting with an immigration-death process over A^* with immigration intensity $\lambda_{\gamma,\delta}^*$, unit per-capita death rate, and equilibrium distribution $\text{PRM}(\lambda_{\gamma,\delta}^*)$. \square

We now again assume that the distributional parameters p_{00n}, q_{1n} and q_{2n} satisfy (4.76) from Section 4.4.4 and apply Theorem 4.4.11 to the case where $A^* = A_n^* = [u_n^*, \infty)^2$. We express the error bounds in terms of the threshold u_n^* and of the probability of simultaneous success p_{11n} .

Proposition 4.4.12. *Let $p_{11n} \in (0, 1)$ and assume that q_{1n}, q_{2n} and p_{00n} satisfy (4.76). For any choice of $u_n^* \geq -\log n$, define $A^* = [u_n^*, \infty)^2$. With the*

notations from Sections 4.4.1-4.4.4,

$$(i) \ d_{TV}(\text{PRM}(\boldsymbol{\lambda}^*), \text{PRM}(\boldsymbol{\lambda}_{\gamma, \delta}^*)) \leq \frac{4(1 + \gamma + \delta)^2 p_{11n}}{[1 - (1 + \gamma + \delta)p_{11n}]^3} e^{-u_n^*},$$

$$(ii) \ d_2(\text{PRM}(\boldsymbol{\lambda}^*), \text{PRM}(\boldsymbol{\lambda}_{\gamma, \delta}^*)) \leq \left(e^{-u_n^*} \wedge 1.65e^{-u_n^*/2} \right) \frac{4(1 + \gamma + \delta)^2 p_{11n}}{[1 - (1 + \gamma + \delta)p_{11n}]^3}.$$

Proof. For ease of notation we again omit the subscript n . (i) By Theorem 4.4.11,

$$\begin{aligned} & d_{TV}(\text{PRM}(\boldsymbol{\lambda}^*), \text{PRM}(\boldsymbol{\lambda}_{\gamma, \delta}^*)) \\ & \leq \int_{u^*}^{\infty} \int_{u^*}^t |\lambda^*(s, t) - \lambda_{\gamma, \delta}^*(s, t)| \, ds dt + \int_{u^*}^{\infty} \int_{u^*}^s |\lambda^*(s, t) - \lambda_{\gamma, \delta}^*(s, t)| \, dt ds \\ & \quad + \int_{u^*}^{\infty} |\dot{\lambda}^*(s) - \dot{\lambda}_{\gamma, \delta}^*(s)| \, ds. \end{aligned}$$

Define

$$h := h(p_{11}) := \frac{1 + \delta}{1 + \gamma + \delta} - \frac{\log q_2}{\log p_{00}} \quad \text{and} \quad g := g(p_{11}) := \frac{1 + \gamma}{1 + \gamma + \delta} - \frac{\log q_1}{\log p_{00}}.$$

We first consider the case $s = t$. Note that, with definitions (4.75) and (4.77),

$$\begin{aligned} \dot{\lambda}^*(s) &= \frac{\log(p_{00}/q_1 q_2)}{\log(1/p_{00})} e^{-s} = \left[\frac{\log q_1 + \log q_2}{\log p_{00}} - \frac{2 + \gamma + \delta}{1 + \gamma + \delta} + \frac{1}{1 + \gamma + \delta} \right] e^{-s} \\ &= \left[\frac{1}{1 + \gamma + \delta} - h(p_{11}) - g(p_{11}) \right] e^{-s}, \end{aligned}$$

and that, since $h, g \geq 0$ by Lemma 4.4.6 (i) and (ii), we thus have $\dot{\lambda}^*(s) \leq \dot{\lambda}_{\gamma, \delta}^*(s)$. Hence,

$$\int_{u^*}^{\infty} |\dot{\lambda}^*(s) - \dot{\lambda}_{\gamma, \delta}^*(s)| \, ds = \int_{u^*}^{\infty} (h + g) e^{-s} \, ds \leq \frac{(\gamma + \delta)p_{11}}{1 - (1 + \gamma + \delta)p_{11}} e^{-u^*},$$

again by Lemma 4.4.6 (i) and (ii). For $s < t$, note that

$$\begin{aligned} \lambda^*(s, t) &= \left[\frac{\gamma}{1 + \gamma + \delta} + h \right] \left[\frac{1 + \delta}{1 + \gamma + \delta} - h \right] e^{-\frac{\gamma}{1 + \gamma + \delta} s} e^{-\frac{1 + \delta}{1 + \gamma + \delta} t} e^{h(t-s)} \\ &= \lambda_{\gamma, \delta}^*(s, t) e^{h(t-s)} + \left(\frac{1 + \delta - \gamma}{1 + \gamma + \delta} h - h^2 \right) e^{-\frac{\gamma}{1 + \gamma + \delta} s} e^{-\frac{1 + \delta}{1 + \gamma + \delta} t} e^{h(t-s)}, \end{aligned}$$

where $\lambda^*(s, t)$ and $\lambda_{\gamma, \delta}^*(s, t)$ are defined by (4.75) and (4.77), respectively. Thereby,

$$\left| \lambda^*(s, t) - \lambda_{\gamma, \delta}^*(s, t) e^{h(t-s)} \right| = \left| \frac{1 + \delta - \gamma}{1 + \gamma + \delta} h - h^2 \right| e^{-\frac{\gamma}{1+\gamma+\delta} s} e^{-\frac{1+\delta}{1+\gamma+\delta} t} e^{h(t-s)},$$

where $\left| \frac{1+\delta-\gamma}{1+\gamma+\delta} h - h^2 \right| \leq h + h^2$. Note that we have

$$\left| \lambda^*(s, t) - \lambda_{\gamma, \delta}^*(s, t) \right| \leq \left| \lambda^*(s, t) - \lambda_{\gamma, \delta}^*(s, t) e^{h(t-s)} \right| + \lambda_{\gamma, \delta}^*(s, t) \left| e^{h(t-s)} - 1 \right|. \quad (4.91)$$

We first compute the following integral:

$$\begin{aligned} \int_{u^*}^{\infty} \int_{u^*}^t e^{-\frac{\gamma}{1+\gamma+\delta} s - \frac{1+\delta}{1+\gamma+\delta} t + h(t-s)} ds dt &= \int_{u^*}^{\infty} \int_{u^*}^t e^{-\frac{\log(p_{00}/q_2)}{\log p_{00}} s - \frac{\log q_2}{\log p_{00}} t} ds dt \\ &= \frac{\log p_{00}}{\log q_2} e^{-u^*}, \end{aligned} \quad (4.92)$$

where, using $z \leq -\log(1-z) \leq \frac{z}{1-z}$ for all $|z| \leq 1$, and (4.76),

$$\begin{aligned} \frac{\log p_{00}}{\log q_2} &= \frac{-\log[1 - (1 + \gamma + \delta)p_{11}]}{-\log[1 - (1 + \delta)p_{11}]} \leq \frac{1 + \gamma + \delta}{(1 + \delta)[1 - (1 + \gamma + \delta)p_{11}]} \\ &\leq \frac{1 + \gamma + \delta}{1 + \delta} \left\{ 1 + \frac{(1 + \gamma + \delta)p_{11}}{[1 - (1 + \gamma + \delta)p_{11}]^2} \right\}. \end{aligned} \quad (4.93)$$

Moreover, by Lemma 4.4.6 (i),

$$h + h^2 \leq \frac{\gamma p_{11}}{1 - (1 + \gamma + \delta)p_{11}} + \left[\frac{\gamma p_{11}}{1 - (1 + \gamma + \delta)p_{11}} \right]^2 \leq \frac{2\gamma p_{11}}{[1 - (1 + \gamma + \delta)p_{11}]^2},$$

since $\gamma p_{11} = p_{10} < 1$, and therefore $(\gamma p_{11})^2 \leq \gamma p_{11}$. Then,

$$\int_{u^*}^{\infty} \int_{u^*}^t \left| \lambda^*(s, t) - \lambda_{\gamma, \delta}^*(s, t) e^{h(t-s)} \right| ds dt \leq \frac{2\gamma(1 + \gamma + \delta)p_{11}}{(1 + \delta)[1 - (1 + \gamma + \delta)p_{11}]^3} e^{-u^*}, \quad (4.94)$$

which gives a bound for the integral of the first error term in (4.91). For the second error term in (4.91), note first that $|e^{h(t-s)} - 1| = e^{h(t-s)} - 1$, since

$h \geq 0$ and $t > s$. By (4.92) and (4.93), and with definition (4.77) of $\lambda_{\gamma,\delta}^*(s, t)$, we obtain

$$\begin{aligned} \int_{u^*}^{\infty} \int_{u^*}^t \lambda_{\gamma,\delta}^*(s, t) e^{h(t-s)} &= \frac{\gamma(1+\delta) \log p_{00}}{(1+\gamma+\delta)^2 \log q_2} e^{-u^*} \\ &\leq \frac{\gamma}{1+\gamma+\delta} \left\{ 1 + \frac{(1+\gamma+\delta)p_{11}}{[1-(1+\gamma+\delta)p_{11}]^2} \right\} e^{-u^*}, \end{aligned} \quad (4.95)$$

whereas

$$\int_{u^*}^{\infty} \int_{u^*}^t \lambda_{\gamma,\delta}^*(s, t) ds dt = \frac{\gamma}{1+\gamma+\delta} e^{-u^*}. \quad (4.96)$$

By (4.95) and (4.96), we may thus bound the integral of the second error term in (4.91) as follows:

$$\int_{u^*}^{\infty} \int_{u^*}^t \lambda_{\gamma,\delta}^*(s, t) \left| e^{h(t-s)} - 1 \right| ds dt \leq \frac{\gamma p_{11}}{[1-(1+\gamma+\delta)p_{11}]^2} e^{-u^*}. \quad (4.97)$$

Hence, for $s < t$, (4.91), (4.94) and (4.97) give

$$\begin{aligned} \int_{u^*}^{\infty} \int_{u^*}^t \left| \lambda^*(s, t) - \lambda_{\gamma,\delta}^*(s, t) \right| ds dt \\ \leq \frac{\gamma p_{11}}{[1-(1+\gamma+\delta)p_{11}]^3} \left\{ \frac{2(1+\gamma+\delta)}{1+\delta} + 1 \right\} e^{-u^*}. \end{aligned}$$

By proceeding analogously for $s > t$, we obtain

$$\begin{aligned} \int_{u^*}^{\infty} \int_{u^*}^s \left| \lambda^*(s, t) - \lambda_{\gamma,\delta}^*(s, t) \right| dt ds \\ \leq \frac{\delta p_{11}}{[1-(1+\gamma+\delta)p_{11}]^3} \left\{ \frac{2(1+\gamma+\delta)}{1+\gamma} + 1 \right\} e^{-u^*}. \end{aligned}$$

The sum of the bounds for the three cases $s = t$, $s < t$ and $s > t$ yields the overall bound

$$\begin{aligned} \int_{A^*} \left| \lambda^*(dz) - \lambda_{\gamma,\delta}^*(dz) \right| \\ \leq \frac{2p_{11}}{[1-(1+\gamma+\delta)p_{11}]^3} \left\{ \frac{\gamma(1+\gamma+\delta)}{1+\delta} + \frac{\delta(1+\gamma+\delta)}{1+\gamma} + \gamma + \delta \right\} e^{-u^*} \\ \leq \frac{4(1+\gamma+\delta)^2 p_{11}}{[1-(1+\gamma+\delta)p_{11}]^3} e^{-u^*}, \end{aligned}$$

where we used $(1 + \gamma)^{-1}, (1 + \delta)^{-1} < 1$, and $\gamma + \delta \leq 1 + \gamma + \delta \leq (1 + \gamma + \delta)^2$ for the second inequality.

(ii) Direct computations give $\lambda^*(A^*) = e^{-u^*} = \lambda_{\gamma, \delta}^*(A^*)$. Theorem 4.4.11 (ii), together with the bound from (i), then immediately gives the result. \square

The error bounds established in Proposition 4.4.12 are similar to the error bound from Proposition 4.4.10. As before, p_{11n} needs to converge to 0 fast enough to make up for the factor $e^{-u_n^*}$ which increases the size of the error as soon as $u_n^* < 0$. And since $u_n^* \geq 0$ gives 1 or no points in A^* , the mean number of points in A^* being given by $e^{-u_n^*}$ for either process, we would certainly want the threshold u_n^* to be negative.

The biggest difference between the d_2 -bounds from Propositions 4.4.10 and 4.4.12 is that the former contains the multiplicative factor $[1 - (1 + \gamma + \delta)p_{11n}]^{-2}$ and the latter the bigger factor $[1 - (1 + \gamma + \delta)p_{11n}]^{-3}$. However, since we need $p_{11n} \rightarrow 0$ as $n \rightarrow \infty$, we will have $(1 + \gamma + \delta)p_{11n} \leq 1/2$ for all n large enough. Then $[1 - (1 + \gamma + \delta)p_{11n}]^{-3} \leq 2[1 - (1 + \gamma + \delta)p_{11n}]^{-2}$ so that both error bounds will be of the same rate. Hence, for large enough n , the approximation by a further Poisson process does not add an error of a bigger size than the one that arises from the approximation by only PRM(λ^*).

4.4.7 Final bound in the d_2 -distance

The following corollary summarises the results from Sections 4.4.2, 4.4.5 and 4.4.6. It gives an estimate for the error in the d_2 -distance of the approximation of the law of an MPPE $\Xi_{A^*}^*$ with i.i.d. Marshall-Olkin geometric marks, living on a lattice of points contained in $A^* \cap [-\log n, \infty)^2$, by the law of a Poisson process with a continuous intensity measure $\lambda_{\gamma, \delta}^*$ over $A^* \cap [-\log n, \infty)^2$, where $A^* = [u^*, \infty)^2$ for some choice of threshold $u^* \geq -\log n$.

Corollary 4.4.13. *Let $p_{11n} \in (0, 1)$ and assume that q_{1n} , q_{2n} and p_{00n} satisfy (4.76). For any choice of $u_n^* \geq -\log n$, define $A^* = [u_n^*, \infty)^2$. With the notations from Sections 4.4.1-4.4.4,*

$$\begin{aligned} & d_2(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\lambda_{\gamma, \delta}^*)) \\ & \leq \frac{e^{-u_n^*}}{n} + \frac{(1 + \gamma + \delta)^2 p_{11n}}{[1 - (1 + \gamma + \delta)p_{11n}]^3} \left\{ 2\sqrt{2} + 7 \left(e^{-u_n^*} \wedge 1.65e^{-u_n^*/2} \right) \right\}. \end{aligned}$$

Proof. We have

$$\begin{aligned} & d_2(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\lambda}_{\gamma, \delta}^*)) \\ & \leq d_2(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\pi}^*)) + d_2(\text{PRM}(\boldsymbol{\pi}^*), \text{PRM}(\boldsymbol{\lambda}^*)) \\ & \quad + d_2(\text{PRM}(\boldsymbol{\lambda}^*), \text{PRM}(\boldsymbol{\lambda}_{\gamma, \delta}^*)). \end{aligned}$$

By (2.42) and Theorem 4.4.1,

$$d_2(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\pi}^*)) \leq d_{TV}(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\pi}^*)) \leq \frac{e^{-u_n^*}}{n}.$$

Furthermore, with the results from Propositions 4.4.10 and 4.4.12, and using $(1 + \gamma + \delta) \leq (1 + \gamma + \delta)^2$ and $[1 - (1 + \gamma + \delta)p_{11}]^{-2} \leq [1 - (1 + \gamma + \delta)p_{11}]^{-3}$, we obtain

$$\begin{aligned} & d_2(\text{PRM}(\boldsymbol{\pi}^*), \text{PRM}(\boldsymbol{\lambda}^*)) + d_2(\text{PRM}(\boldsymbol{\lambda}^*), \text{PRM}(\boldsymbol{\lambda}_{\gamma, \delta}^*)) \\ & \leq \frac{(1 + \gamma + \delta)^2 p_{11n}}{[1 - (1 + \gamma + \delta)p_{11n}]^3} \left\{ 2\sqrt{2} + 7 \left(e^{-u_n^*} \wedge 1.65e^{-u_n^*/2} \right) \right\}. \end{aligned}$$

□

By far the smallest component of the error estimate from Corollary 4.4.13 is given by $e^{-u_n^*}/n$, the error arising from approximating $\mathcal{L}(\Xi_{A^*}^*)$ by $\text{PRM}(\mathbb{E}\Xi_{A^*}^*)$, which lives on the lattice $A^* \cap E^*$ just as $\Xi_{A^*}^*$. This part of the error corresponds exactly to the overall error estimate that we obtained in Section 4.2.3 for the approximation of an MPPE $\Xi_{A^*}^*$ with Marshall-Olkin exponential marks by a Poisson process with mean measure $\mathbb{E}\Xi_{A^*}^*$. Yet the Marshall-Olkin exponential is a continuous distribution and the mean measure $\mathbb{E}(\Xi_{A^*}^*)$ is thereby also continuous. As for MPPE's with univariate geometric marks (see Section 3.2.2), a far bigger error emerges for the MPPE with Marshall-Olkin geometric marks when going from the Poisson process on the lattice to a Poisson process on $A^* \cap [-\log n, \infty)^2$ with continuous intensity. This error can only be small if the probability of simultaneous success of the Marshall-Olkin geometric distribution, p_{11} , and thereby also the marginal success probabilities $1 - q_{1n}$ and $1 - q_{2n}$, tend to zero as $n \rightarrow \infty$ at a rate fast enough to compensate for the factor $e^{-u_n^*}$, the (rough) number of points expected in A_n^* for each of the processes. For instance, for $A_n^* = [-\log \log n, \infty)^2$ and $p_{11n} = 1/n$, we expect $\log n$

joint threshold exceedances, and obtain

$$\begin{aligned}
 & d_2(\mathcal{L}(\Xi_{A^*}^*), \text{PRM}(\boldsymbol{\lambda}_{\gamma, \delta}^*)) \\
 & \leq \frac{\log n}{n} + \frac{(1 + \gamma + \delta)^2}{n[1 - (1 + \gamma + \delta)/n]^3} \left\{ 2\sqrt{2} + 7 \left(\log n \wedge 1.65\sqrt{\log n} \right) \right\} \\
 & \leq \frac{C \log n}{n},
 \end{aligned}$$

where C is some constant. With the (very strong) condition $p_{11n} = 1/n$, we thus obtain an error of the same size as the error that we obtain when approximating $\mathcal{L}(\Xi_{A^*}^*)$ only by $\text{PRM}(\mathbb{E}\Xi_{A^*}^*)$.

Bibliography

- Ambrosio, L., Gigli, N., and Savaré, G. (2005). *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel.
- Anderson, C. W., Coles, S. G., and Hüsler, J. (1997). Maxima of Poisson-like variables and related triangular arrays. *Ann. Appl. Probab.*, 7:953–971.
- Barbour, A. D. (1988). Stein’s method and Poisson process convergence. *J. Appl. Probab.*, (Special Vol. 25A):175–184. A celebration of applied probability.
- Barbour, A. D. and Brown, T. C. (1992). Stein’s method and point process approximation. *Stochastic Process. Appl.*, 43(1):9–31.
- Barbour, A. D. and Eagleson, G. K. (1983). Poisson approximation for some statistics based on exchangeable trials. *Adv. in Appl. Probab.*, 15(3):585–600.
- Barbour, A. D. and Hall, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.*, 95(3):473–480.
- Barbour, A. D., Holst, L., and Janson, S. (1992). *Poisson approximation*. The Clarendon Press Oxford University Press.
- Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.
- Charpentier, A. and Segers, J. (2009). Tails of multivariate archimedean copulas. *Journal of Multivariate Analysis*, 100:1521–1537.
- Chen, L. H. Y. (1975a). An approximation theorem for sums of certain randomly selected indicators. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 33(1):69–74.

- Chen, L. H. Y. (1975b). Poisson approximation for dependent trials. *Ann. Probability*, 3(3):534–545.
- Chen, L. H. Y. (1998). Stein’s method: some perspectives with applications. In Accardi, L. and Heyde, C. C., editors, *Probability Towards 2000*, volume 128 of *Lecture Notes in Statistics*, pages 97–122. Springer, New York.
- Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Vol. I: Elementary Theory and Methods*. Probability and its Applications. Springer, New York, 2nd edition.
- Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes, Vol. II: General Theory and Structure*. Probability and its Applications. Springer, New York, 2nd edition.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, New York.
- Erhardsson, T. (2005). Stein’s method for Poisson and compound Poisson approximation. In Barbour, A. D. and Chen, L. H. Y., editors, *An Introduction to Stein’s method*, volume 4 of *Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore*, pages 61–113. Singapore University Press, Singapore.
- Feidt, A., Genest, C., and Nešlehová, J. (2010). Asymptotics of joint maxima for discontinuous random variables. *Extremes*, 13:35–53.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190.
- Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *Astin Bull.*, 37:475–515.
- Gumbel, E. J. (1958). Distributions à plusieurs variables dont les marges sont données. *C. R. Acad. Sci. Paris*, 246:2717–2719.
- Gumbel, E. J. (1965). Two systems of bivariate extremal distributions. *Bull. Inst. Internat. Statist.*, 41:749–763.
- Hall, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.*, 16(2):433–439.

- Hawkes, A. G. (1972). A bivariate exponential distribution with applications to reliability. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(1):pp. 129–131.
- Joe, H. (1993). Multivariate dependence measures and data analysis. *Comput. Statist. Data Anal.*, 16:279–297.
- Joe, H. (1997). *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.
- Kallenberg, O. (1986). *Random measures*. Akademie-Verlag, Berlin, 4th edition.
- Kerstan, J. (1964). Verallgemeinerung eines Satzes von Prochorow und Le Cam. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:173–179.
- Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions*, Second Edition. Wiley, New York.
- Le Cam, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.*, 10:1181–1197.
- Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.
- Lindvall, T. (2002). *Lectures on the coupling method*. Dover Publications Inc., Mineola, NY. Corrected reprint of the 1992 original.
- Marshall, A. W. and Olkin, I. (1967a). A generalized bivariate exponential distribution. *J. Appl. Probability*, 4:291–302.
- Marshall, A. W. and Olkin, I. (1967b). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):pp. 30–44.
- Marshall, A. W. and Olkin, I. (1985). A family of bivariate distributions generated by the bivariate Bernoulli distribution. *J. Amer. Statist. Assoc.*, 80:332–338.
- McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, Princeton, NJ.
- Michel, R. (1988). An improved error bound for the compound poisson approximation of a nearly homogeneous portfolio. *Astin Bull.*, 17:165–169.

- Mitov, K. and Nadarajah, S. (2005). Limit distributions for the bivariate geometric maxima. *Extremes*, 8:357–370.
- Nadarajah, S. and Mitov, K. (2002). Asymptotics of maxima of discrete random variables. *Extremes*, 5:287–294.
- Nelsen, R. B. (2006). *An Introduction to Copulas*, Second Edition. Springer, New York.
- Preston, C. (1975). Spatial birth-and-death processes. *Bull. ISI*, 46(2):371–391.
- Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
- Serfling, R. J. (1975). A general Poisson approximation theorem. *Annals of Probability*, 3(4):726–731.
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.
- Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602, Berkeley, Calif. Univ. California Press.
- Stein, C. (1986). *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA.
- Thorisson, H. (2000). *Coupling, stationarity, and regeneration*. Probability and its Applications. Springer, New York.
- Xia, A. (2005). Stein’s method and Poisson process approximation. In Barbour, A. D. and Chen, L. H. Y., editors, *An Introduction to Stein’s method*, volume 4 of *Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore*, pages 115–181. Singapore University Press, Singapore.